

::ODMA\MHODMA\manage;123935;1
DMH/JCH/(MKM)/dmr
March 29, 2000

PATENT APPLICATION
Docket No.: 2709.1001001

BOX 1 OF 3	Date: <u>4/4/00</u>	Express Mail Label No. <u>EL290720882 US</u> Pages <u>Specification pages 1-1352; Sequence Listings pages 1-1600 and CD</u>
BOX 2 OF 3	Date: <u>4/4/00</u>	Express Mail Label No. <u>EL290720896 US</u> Pages <u>Sequence Listings, pages 1001-3500</u>
BOX 3 OF 3	Date: <u>4/4/00</u>	Express Mail Label No. <u>EL290720905 US</u> Pages <u>Sequence Listings, pages 3501-6222</u>

TITLE OF THE INVENTION

NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO *BACTEROIDES FRAGILIS* FOR DIAGNOSTICS AND THERAPEUTICS

INVENTOR: Gary L. Breton

RELATED APPLICATIONS:

This application claims the benefit of U.S. Provisional Application Serial Number 60/128,705, filed April 9, 1999, the entire teachings of which are incorporated herein by reference.

FIELD OF THE INVENTION

The invention relates to isolated nucleic acids and polypeptides derived from *Bacteroides fragilis* that are useful as molecular targets for diagnostics, prophylaxis and treatment of pathological conditions, as well as materials and methods for the diagnosis, prevention, and amelioration of pathological conditions resulting from bacterial infection.

BACKGROUND OF THE INVENTION

The genus *Bacteroides* is a member of the family *Bacteroidaceae*. They are Gram-negative, obligately anaerobic, nonsporeforming rods. The genus contains at least 39 species, and are often isolated from sewage as well as the digestive tract of man, animals, and insects. *Bacteroides fragilis* was first described in 1898 by Veillon and Zuber, but was called *Bacillus fragilis*. In 1919, Castellani and Chalmers transferred it to the *Bacteroides* genus. The "*B. fragilis* group" refers to the saccharoclastic bacteroids that grow well in bile. Members of this group were previously subspecies of *B. fragilis* and include *B. fragilis*, *B. distasonis*, *B. ovatus*, *B. thetaiotaomicron*, and *B. vulgatus* (Castellani and Chalmers. 1984. Genus I. *Bacteroides* 1919, 959. Krieg and Holt (editors) In Bergey's Manual of Systematic Bacteriology, 1:604-631).

Bacteroides fragilis accounts for only 1% of the normal flora of the human colon, but is the most common anaerobe isolated from clinical specimens. It is associated with soft tissue infections, abscesses and bacteremia (Moncrief J., *et al*, 1998. Infect. Immun. 66:1735-1739). *B. fragilis* has also been associated with infection of the skeletal muscle (Katagiri, K., *et al*, 1996. J. Dermatology. 23:129-132), and meningitis (Aucher, P., *et al*, 1996. Eur. J. Clin. Microbiol. Infect. Dis. 15:820-823). The *B. fragilis* group is responsible for 65% of all anaerobic bacteremia cases, with mortality rates in excess of 19% (Redondo, M., *et al*, 1995. Clinical Infectious Disease. 20:1492-1496).

In 1984, strains of *B. fragilis* were found to cause diarrhea in newborn lambs (Myers, L., *et al*, 1984. Infect. Immun. 44:241-244). Subsequently, it has been shown that *B. fragilis* is associated with diarrhea in other livestock and young children. These strains are called enterotoxigenic strains, because they produced a 20KD metalloprotease enterotoxin with intestinal secretory activity (Moncrief J., *et al*, 1995. Infect. Immun. 63:175-181).

There has been an increase in antibiotic resistance within the *Bacteroides fragilis* group. While there is still excellent activity of many antibiotics, even some of the most potent agents, the carbapenems and the β -lactamase-inhibitor combinations, are losing

activity (Snydman,D., *et al*, 1996. Clinical Infectious Diseases. 23:S54-65). The cefoxitin resistance rate has increased from 0% in 1987 to 22% in 1995 (Bianchini, H., *et al*, 1997. Clinical Infectious Diseases. 25:S268-269). Resistance to metronidazole, co-amoxiclav, and imipenem is rare, but strains have been found that are resistant to one or
 5 all of these antibiotics. (Turner,P., *et al*, 1995.The Lancet. 345:1275-1277). Clindaycin resistance has been shown to be transferred between strains by either plasmid or transposon mechanisms. (Dalmau, D., *et al*, 1997. Clinical Infectious Diseases. 24:874-877). The increasing resistance to antibiotics commonly used against *Bacteroides* species may eventually lead to failures of these treatments.

10 Sequencing and analysis of this genome is crucial for the identification of essential genes for development of drug targets and to reduce the emerging health threat this organism poses.

SUMMARY OF THE INVENTION

15 The present invention fulfills the need for diagnostic tools and therapeutics by providing bacterial-specific compositions and methods for detecting *Bacteroides* species including *B. fragilis* , as well as compositions and methods useful for treating and preventing *Bacteroides* infection, in particular, *B. fragilis* infection, in vertebrates including mammals.

20 The present invention encompasses isolated nucleic acids and polypeptides derived from *B. fragilis* that are useful as reagents for diagnosis of bacterial disease, components of effective antibacterial vaccines, and/or as targets for antibacterial drugs including anti-*B. fragilis* drugs. They can also be used to detect the presence of *B. fragilis* and other *Bacteroides* species in a sample; and in screening compounds for the
 25 ability to interfere with the *B. fragilis* life cycle or to inhibit *B. fragilis* infection. They also have use as biocontrol agents for plants.

In one aspect, the invention features compositions of nucleic acids corresponding to entire coding sequences of *B. fragilis* proteins, including surface or secreted proteins or parts thereof, nucleic acids capable of binding mRNA from *B. fragilis* proteins to

block protein translation, and methods for producing *B. fragilis* proteins or parts thereof using peptide synthesis and recombinant DNA techniques. This invention also features antibodies and nucleic acids useful as probes to detect *B. fragilis* infection. In addition, vaccine compositions and methods for the protection or treatment of infection by *B.*

5 *fragilis* are within the scope of this invention.

The nucleotide sequences provided in SEQ ID NO: 1 - SEQ ID NO: 5222, a fragment thereof, or a nucleotide sequence at least about 99.5% identical to a sequence contained within SEQ ID NO: 1 - SEQ ID NO: 5222 may be "provided" in a variety of medias to facilitate use thereof. As used herein, "provided" refers to a manufacture,
10 other than an isolated nucleic acid molecule, which contains a nucleotide sequence of the present invention, i.e., the nucleotide sequence provided in SEQ ID NO: 1 - SEQ ID NO: 5222, a fragment thereof, or a nucleotide sequence at least about 99.5% identical to a sequence contained within SEQ ID NO: 1 - SEQ ID NO: 5222. Uses for and methods for providing nucleotide sequences in a variety of media is well known in the art (see e.g.,
15 EPO Publication No. EP 0 756 006).

In one application of this embodiment, a nucleotide sequence of the present invention can be recorded on computer readable media. As used herein, "computer readable media" refers to any media which can be read and accessed directly by a computer. Such media include, but are not limited to: magnetic storage media, such as
20 floppy discs, hard disc storage media, and magnetic tape; optical storage media such as CD-ROM; electrical storage media such as RAM and ROM; and hybrids of these categories such as magnetic/optical storage media. A person skilled in the art can readily appreciate how any of the presently known computer readable media can be used to create a manufacture comprising computer readable media having recorded thereon a
25 nucleotide sequence of the present invention.

As used herein, "recorded" refers to a process for storing information on computer readable media. A person skilled in the art can readily adopt any of the presently known methods for recording information on computer readable media to

generate manufactures comprising the nucleotide sequence information of the present invention.

A variety of data storage structures are available to a person skilled in the art for creating a computer readable media having recorded thereon a nucleotide sequence of the present invention. The choice of the data storage structure will generally be based on the means chosen to access the stored information. In addition, a variety of data processor programs and formats can be used to store the nucleotide sequence information of the present invention on computer readable media. The sequence information can be represented in a word processing text file, formatted in commercially-available software such as WordPerfect and Microsoft Word, or represented in the form of an ASCII file, stored in a database application, such as DB2, Sybase, Oracle, or the like. A person skilled in the art can readily adapt any number of data processor structuring formats (e.g. text file or database) in order to obtain computer readable media having recorded thereon the nucleotide sequence information of the present invention.

By providing the nucleotide sequence of SEQ ID NO: 1 - SEQ ID NO: 5222, a fragment thereof, or a nucleotide sequence at least about 99.5% identical to SEQ ID NO: 1 - SEQ ID NO: 5222 in computer readable form, a person skilled in the art can routinely access the coding sequence information for a variety of purposes. Computer software is publicly available which allows a person skilled in the art to access sequence information provided in a computer readable media. Examples of such computer software include programs of the "Staden Package", "DNA Star", "MacVector", GCG "Wisconsin Package" (Genetics Computer Group, Madison, WI) and "NCBI Toolbox" (National Center For Biotechnology Information). Suitable programs are described, for example, in Martin J. Bishop, ed., *Guide to Human Genome Computing*, 2d Edition, Academic Press, San Diego, CA. (1998); and Leonard F. Peruski, Jr., and Anne Harwood Peruski, *The Internet and the New Biology: Tools for Genomic and Molecular Research*, American Society for Microbiology, Washington, D.C. (1997).

Computer algorithms enable the identification of *B. fragilis* open reading frames (ORFs) within SEQ ID NO: 1 - SEQ ID NO: 5222 which contain homology to ORFs or proteins from other organisms. Examples of such similarity-search algorithms include the BLAST [Altschul et al., J. Mol. Biol. 215:403-410 (1990)] and Smith-Waterman
5 [Smith and Waterman (1981) Advances in Applied Mathematics, 2:482-489] search algorithms. Suitable search algorithms are described, for example, in Martin J. Bishop, ed., *Guide to Human Genome Computing*, 2d Edition, Academic Press, San Diego, CA. (1998); and Leonard F. Peruski, Jr., and Anne Harwood Peruski, *The Internet and the New Biology: Tools for Genomic and Molecular Research*, American Society for
10 Microbiology, Washington, D.C. (1997). Such algorithms are utilized on computer systems as exemplified below. The ORFs so identified represent protein encoding fragments within the *B. fragilis* genome and are useful in producing commercially important proteins such as enzymes used in fermentation reactions and in the production of commercially useful metabolites.

15 The present invention further provides systems, particularly computer-based systems, which contain the sequence information described herein. Such systems are designed to identify commercially important fragments of the *B. fragilis* genome. As used herein, "a computer-based system" refers to the hardware means, software means, and data storage means used to analyze the nucleotide sequence information of the
20 present invention. The minimum hardware means of the computer-based systems of the present invention comprises a central processing unit (CPU), input means, output means, and data storage means. A person skilled in the art can readily appreciate that any one of the currently available computer-based systems is suitable for use in the present invention. The computer-based systems of the present invention comprise a data storage
25 means having stored therein a nucleotide sequence of the present invention and the necessary hardware means and software means for supporting and implementing a search means. As used herein, "data storage means" refers to memory which can store nucleotide sequence information of the present invention, or a memory access means

which can access manufactures having recorded thereon the nucleotide sequence information of the present invention.

As used herein, "search means" refers to one or more programs which are implemented on the computer-based system to compare a target sequence or target structural motif with the sequence information stored within the data storage means. Search means are used to identify fragments or regions of the *B. fragilis* genome which are similar to, or "match", a particular target sequence or target motif. A variety of known algorithms are known in the art and have been disclosed publicly, and a variety of commercially available software for conducting homology-based similarity searches are available and can be used in the computer-based systems of the present invention. Examples of such software includes, but is not limited to, FASTA (GCG Wisconsin Package), Bic_SW (Compugen Bioccelerator), BLASTN2, BLASTP2, BLASTX2 (NCBI) and Motifs (GCG). Suitable software programs are described, for example, in Martin J. Bishop, ed., *Guide to Human Genome Computing*, 2d Edition, Academic Press, San Diego, CA. (1998); and Leonard F. Peruski, Jr., and Anne Harwood Peruski, *The Internet and the New Biology: Tools for Genomic and Molecular Research*, American Society for Microbiology, Washington, D.C. (1997). A person skilled in the art can readily recognize that any one of the available algorithms or implementing software packages for conducting homology searches can be adapted for use in the present computer-based systems.

As used herein, a "target sequence" can be any DNA or amino acid sequence of six or more nucleotides or two or more amino acids. A person skilled in the art can readily recognize that the longer a target sequence is, the less likely a target sequence will be present as a random occurrence in the database. The most preferred sequence length of a target sequence is from about 10 to 100 amino acids or from about 30 to 300 nucleotide residues. However, it is well recognized that many genes are longer than 500 amino acids, or 1.5 kb in length, and that commercially important fragments of the *B. fragilis*

genome, such as sequence fragments involved in gene expression and protein processing, will often be shorter than 30 nucleotides.

As used herein, "a target structural motif," or "target motif," refers to any rationally selected sequence or combination of sequences in which the sequence(s) are
5 chosen based on a specific functional domain or three-dimensional configuration which is formed upon the folding of the target polypeptide. There are a variety of target motifs known in the art. Protein target motifs include, but are not limited to, enzymatic active sites, membrane-spanning regions, and signal sequences. Nucleic acid target motifs include, but are not limited to, promoter sequences, hairpin structures and inducible
10 expression elements (protein binding sequences).

A variety of structural formats for the input and output means can be used to input and output the information in the computer-based systems of the present invention. A preferred format for an output means ranks fragments of the *B. fragilis* genome possessing varying degrees of homology to the target sequence or target motif. Such
15 presentation provides a person skilled in the art with a ranking of sequences which contain various amounts of the target sequence or target motif and identifies the degree of homology contained in the identified fragment.

A variety of comparing means can be used to compare a target sequence or target motif with the data storage means to identify sequence fragments of the *B. fragilis*
20 genome. In the present examples, implementing software which implement the BLASTP2 and bic_SW algorithms (Altschul et al., J Mol. Biol. 215:403-410 (1990); Compugen Biocellator) was used to identify open reading frames within the *B. fragilis* genome. A person skilled in the art can readily recognize that any one of the publicly available homology search programs can be used as the search means for the computer-
25 based systems of the present invention. Suitable programs are described, for example, in Martin J. Bishop, ed., *Guide to Human Genome Computing*, 2d Edition, Academic Press, San Diego, CA. (1998); and Leonard F. Peruski, Jr., and Anne Harwood Peruski, *The*

Internet and the New Biology: Tools for Genomic and Molecular Research, American Society for Microbiology, Washington, D.C. (1997).

The invention features *B. fragilis* polypeptides, preferably a substantially pure preparation of an *B. fragilis* polypeptide, or a recombinant *B. fragilis* polypeptide. In preferred embodiments: the polypeptide has biological activity; the polypeptide has an amino acid sequence at least about 60%, 70%, 80%, 90%, 95%, 98%, or 99% identical to an amino acid sequence of the invention contained in the Sequence Listing, preferably it has about 65% sequence identity with an amino acid sequence of the invention contained in the Sequence Listing, and most preferably it has about 92% to about 99% sequence identity with an amino acid sequence of the invention contained in the Sequence Listing; the polypeptide has an amino acid sequence essentially the same as an amino acid sequence of the invention contained in the Sequence Listing; the polypeptide is at least about 5, 10, 20, 50, 100, or 150 amino acid residues in length; the polypeptide includes at least about 5, preferably at least about 10, more preferably at least about 20, still more preferably at least about 50, 100, or 150 contiguous amino acid residues of the invention contained in the Sequence Listing. In yet another preferred embodiment, the amino acid sequence which differs in sequence identity by about 7% to about 8% from the *B. fragilis* amino acid sequences of the invention contained in the Sequence Listing is also encompassed by the invention.

In preferred embodiments: the *B. fragilis* polypeptide is encoded by a nucleic acid of the invention contained in the Sequence Listing, or by a nucleic acid having at least about 60%, 70%, 80%, 90%, 95%, 98%, or 99% homology with a nucleic acid of the invention contained in the Sequence Listing.

In a preferred embodiment, the subject *B. fragilis* polypeptide differs in amino acid sequence at about 1, 2, 3, 5, 10 or more residues from a sequence of the invention contained in the Sequence Listing. The differences, however, are such that the *B. fragilis* polypeptide exhibits an *B. fragilis* biological activity, e.g., the *B. fragilis* polypeptide retains a biological activity of a naturally occurring *B. fragilis* enzyme.

In preferred embodiments, the polypeptide includes all or a fragment of an amino acid sequence of the invention contained in the Sequence Listing; fused, in reading frame, to additional amino acid residues, preferably to residues encoded by genomic DNA 5' or 3' to the genomic DNA which encodes a sequence of the invention contained
5 in the Sequence Listing.

In yet other preferred embodiments, the *B. fragilis* polypeptide is a recombinant fusion protein having a first *B. fragilis* polypeptide portion and a second polypeptide portion, e.g., a second polypeptide portion having an amino acid sequence unrelated to *B. fragilis*. The second polypeptide portion can be, e.g., any of glutathione-S-transferase, a
10 DNA binding domain, or a polymerase activating domain. In preferred embodiment the fusion protein can be used in a two-hybrid assay.

Polypeptides of the invention include those which arise as a result of alternative transcription events, alternative RNA splicing events, and alternative translational and postranslational events.

15 In a preferred embodiment, the encoded *B. fragilis* polypeptide differs (e.g., by amino acid substitution, addition or deletion of at least one amino acid residue) in amino acid sequence at about 1, 2, 3, 5, 10 or more residues, from a sequence of the invention contained in the Sequence Listing. The differences, however, are such that: the *B. fragilis* encoded polypeptide exhibits an *B. fragilis* biological activity, e.g., the encoded
20 *B. fragilis* enzyme retains a biological activity of a naturally occurring *B. fragilis*.

In preferred embodiments, the encoded polypeptide includes all or a fragment of an amino acid sequence of the invention contained in the Sequence Listing; fused, in reading frame, to additional amino acid residues, preferably to residues encoded by genomic DNA 5' or 3' to the genomic DNA which encodes a sequence of the invention
25 contained in the Sequence Listing.

The *B. fragilis* strain, 14062, from which genomic sequences have been sequenced, has been deposited on July 20, 1998, in the American Type Culture Collection and assigned the ATCC designation # 202158.

Included in the invention are: allelic variations; natural mutants; induced mutants; proteins encoded by DNA that hybridize under high or low stringency conditions to a nucleic acid which encodes a polypeptide of the invention contained in the Sequence Listing (for definitions of high and low stringency see Current Protocols in Molecular Biology, John Wiley & Sons, New York, 1989, 6.3.1 - 6.3.6, hereby incorporated by reference); and, polypeptides specifically bound by antisera to *B. fragilis* polypeptides, especially by antisera to an active site or binding domain of *B. fragilis* polypeptide. The invention also includes fragments, preferably biologically active fragments. These and other polypeptides are also referred to herein as *B. fragilis* polypeptide analogs or variants.

The invention further provides nucleic acids, e.g., RNA or DNA and their respective complements, encoding a polypeptide of the invention. This includes double stranded nucleic acids as well as coding and antisense single strands.

In preferred embodiments, the subject *B. fragilis* nucleic acid will include a transcriptional regulatory sequence, e.g., at least one of a transcriptional promoter or transcriptional enhancer sequence, operably linked to the *B. fragilis* gene sequence, e.g., to render the *B. fragilis* gene sequence suitable for expression in a recombinant host cell.

In yet a further preferred embodiment, the nucleic acid which encodes an *B. fragilis* polypeptide of the invention, hybridizes under stringent conditions to a nucleic acid probe corresponding to at least about 8 consecutive nucleotides of the invention contained in the Sequence Listing; more preferably to at least about 12 consecutive nucleotides of the invention contained in the Sequence Listing; still more preferably to at least about 20 consecutive nucleotides of the invention contained in the Sequence Listing; most preferably to at least about 40 consecutive nucleotides of the invention contained in the Sequence Listing.

In another aspect, the invention provides a substantially pure nucleic acid having a nucleotide sequence which encodes an *B. fragilis* polypeptide. In preferred embodiments: the encoded polypeptide has biological activity; the encoded polypeptide

has an amino acid sequence at least about 60%, 70%, 80%, 90%, 95%, 98% or 99% homologous to an amino acid sequence of the invention contained in the Sequence Listing; the encoded polypeptide has an amino acid sequence essentially the same as an amino acid sequence of the invention contained in the Sequence Listing; the encoded
5 polypeptide is at least about 5, 10, 20, 50, 100, or 150 amino acids in length; the encoded polypeptide comprises at least about 5, preferably at least about 10, more preferably at least about 20, still more preferably at least about 50, 100, or 150 contiguous amino acids of the invention contained in the Sequence Listing.

In another aspect, the invention encompasses: a vector including a nucleic acid
10 which encodes an *B. fragilis* polypeptide or an *B. fragilis* polypeptide variant as described herein; a host cell transfected with the vector; and a method of producing a recombinant *B. fragilis* polypeptide or *B. fragilis* polypeptide variant; including culturing the cell, e.g., in a cell culture medium, and isolating an *B. fragilis* or *B. fragilis* polypeptide variant, e.g., from the cell or from the cell culture medium.

One embodiment of the invention is directed to substantially isolated nucleic
15 acids. Nucleic acids of the invention include sequences comprising at least about 8 nucleotides in length, more preferably at least about 12 nucleotides in length, even more preferably at least about 15-20 nucleotides in length, that correspond to a subsequence of any one of SEQ ID NO: 1 - SEQ ID NO: 5222 or complements thereof. Alternatively,
20 the nucleic acids comprise sequences contained within any ORF (open reading frame), including a complete protein-coding sequence, of which any of SEQ ID NO: 1 - SEQ ID NO: 5222 forms a part. The invention encompasses sequence-conservative variants and function-conservative variants of these sequences. The nucleic acids may be DNA, RNA, DNA/RNA duplexes, protein-nucleic acid (PNA), or derivatives thereof.

25 In another aspect, the invention features a purified recombinant nucleic acid having at least about 50%, 60%, 70%, 80%, 90%, 95%, 98%, or 99% sequence identity or % homology with a sequence of the invention contained in the Sequence Listing

The invention also encompasses recombinant DNA (including DNA cloning and expression vectors) comprising these *B. fragilis* -derived sequences; host cells comprising such DNA, including fungal, bacterial, yeast, plant, insect, and mammalian host cells; and methods for producing expression products comprising RNA and

5 polypeptides encoded by the *B. fragilis* sequences. These methods are carried out by incubating a host cell comprising an *B. fragilis* -derived nucleic acid sequence under conditions in which the sequence is expressed. The host cell may be native or recombinant. The polypeptides can be obtained by (a) harvesting the incubated cells to produce a cell fraction and a medium fraction; and (b) recovering the *B. fragilis*

10 polypeptide from the cell fraction, the medium fraction, or both. The polypeptides can also be made by *in vitro* translation.

In another aspect, the invention features nucleic acids capable of binding mRNA of *B. fragilis* . Such nucleic acid is capable of acting as antisense nucleic acid to control the translation of mRNA of *B. fragilis* . A further aspect features a nucleic acid which is

15 capable of binding specifically to an *B. fragilis* nucleic acid. These nucleic acids are also referred to herein as complements and have utility as probes and as capture reagents.

In another aspect, the invention features an expression system comprising an open reading frame corresponding to *B. fragilis* nucleic acid. The nucleic acid further comprises a control sequence compatible with an intended host. The expression system

20 is useful for making polypeptides corresponding to *B. fragilis* nucleic acid.

In another aspect, the invention encompasses: a vector including a nucleic acid which encodes an *B. fragilis* polypeptide or an *B. fragilis* polypeptide variant as described herein; a host cell transfected with the vector; and a method of producing a recombinant *B. fragilis* polypeptide or *B. fragilis* polypeptide variant; including culturing

25 the cell, e.g., in a cell culture medium, and isolating the *B. fragilis* or *B. fragilis* polypeptide variant, e.g., from the cell or from the cell culture medium.

In yet another embodiment of the invention encompasses reagents for detecting bacterial infection, including *B. fragilis* infection, which comprise at least one *B. fragilis*

-derived nucleic acid defined by any one of SEQ ID NO: 1 - SEQ ID NO: 5222, or sequence-conservative or function-conservative variants thereof. Alternatively, the diagnostic reagents comprise nucleotide sequences that are contained within any open reading frames (ORFs), including preferably complete protein-coding sequences,
 5 contained within any of SEQ ID NO: 1 - SEQ ID NO: 5222, or polypeptide sequences contained within any of SEQ ID NO: 5223 - SEQ ID NO: 10444, or polypeptides of which any of the above sequences forms a part, or antibodies directed against any of the above peptide sequences or function-conservative variants and/or fragments thereof.

The invention further provides antibodies, preferably monoclonal antibodies,
 10 which specifically bind to the polypeptides of the invention. Methods are also provided for producing antibodies in a host animal. The methods of the invention comprise immunizing an animal with at least one *B. fragilis* -derived immunogenic component, wherein the immunogenic component comprises one or more of the polypeptides encoded by any one of SEQ ID NO: 1 - SEQ ID NO: 5222 or sequence-conservative or
 15 function-conservative variants thereof; or polypeptides that are contained within any ORFs, including complete protein-coding sequences, of which any of SEQ ID NO: 1 - SEQ ID NO: 5222 forms a part; or polypeptide sequences contained within any of SEQ ID NO: 5223 - SEQ ID NO: 10444; or polypeptides of which any of SEQ ID NO: 5223 - SEQ ID NO: 10444 forms a part. Host animals include any warm blooded animal,
 20 including without limitation mammals and birds. Such antibodies have utility as reagents for immunoassays to evaluate the abundance and distribution of *B. fragilis* -specific antigens.

In yet another aspect, the invention provides diagnostic methods for detecting *B. fragilis* antigenic components or anti-*B. fragilis* antibodies in a sample. *B. fragilis*
 25 antigenic components may be detected by known processes, including but not limited to detection by a process comprising: (i) contacting a sample suspected to contain a bacterial antigenic component with a bacterial-specific antibody, under conditions in which a stable antigen-antibody complex can form between the antibody and bacterial

antigenic components in the sample; and (ii) detecting any antigen-antibody complex formed in step (i), wherein detection of an antigen-antibody complex indicates the presence of at least one bacterial antigenic component in the sample. In different embodiments of this method, the antibodies used are directed against a sequence encoded
 5 by any of SEQ ID NO: 1 - SEQ ID NO: 5222 or sequence-conservative or function-conservative variants thereof, or against a polypeptide sequence contained in any of SEQ ID NO: 5223 - SEQ ID NO: 10444 or function-conservative variants thereof.

In yet another aspect, the invention provides a method for detecting antibacterial-specific antibodies in a sample, which comprises: (i) contacting a sample suspected to
 10 contain antibacterial-specific antibodies with an *B. fragilis* antigenic component, under conditions in which a stable antigen-antibody complex can form between the *B. fragilis* antigenic component and antibacterial antibodies in the sample; and (ii) detecting any antigen-antibody complex formed in step (i), wherein detection of an antigen-antibody complex indicates the presence of antibacterial antibodies in the sample. In different
 15 embodiments of this method, the antigenic component is encoded by a sequence contained in any of SEQ ID NO: 1 - SEQ ID NO: 5222 or sequence-conservative and function-conservative variants thereof, or is a polypeptide sequence contained in any of SEQ ID NO: 5223 - SEQ ID NO: 10444 or function-conservative variants thereof.

In another aspect, the invention features a method of generating vaccines for
 20 immunizing an individual against *B. fragilis*. The method includes: immunizing a subject with an *B. fragilis* polypeptide, e.g., a surface or secreted polypeptide, or a combination of such peptides or active portion(s) thereof, and a pharmaceutically acceptable carrier. Such vaccines have therapeutic and prophylactic utilities.

In another aspect, the invention features a method of evaluating a compound, e.g.,
 25 a polypeptide, e.g., a fragment of a host cell polypeptide, for the ability to bind an *B. fragilis* polypeptide. The method includes contacting the compound to be evaluated with an *B. fragilis* polypeptide and determining if the compound binds or otherwise interacts with the *B. fragilis* polypeptide. Compounds which bind or otherwise interact with *B.*

fragilis polypeptides are candidates as modulators, including activators and inhibitors, of the bacterial life cycle. These assays can be performed *in vitro* or *in vivo*.

In another aspect, the invention features a method of evaluating a compound, e.g., a polypeptide, e.g., a fragment of a host cell polypeptide, for the ability to bind an *B.*

5 *fragilis* nucleic acid, e.g., DNA or RNA. The method includes contacting the compound to be evaluated with an *B. fragilis* nucleic acid and determining if the compound binds or otherwise interacts with the *B. fragilis* nucleic acid. Compounds which bind *B. fragilis* are candidates as modulators, including activators and inhibitors, of the bacterial life cycle. These assays can be performed *in vitro* or *in vivo*.

10 A particularly preferred embodiment of the invention is directed to a method of screening test compounds for anti-bacterial activity, which method comprises: selecting as a target a bacterial specific sequence, which sequence is essential to the viability of a bacterial species; contacting a test compound with said target sequence; and selecting those test compounds which bind to said target sequence as potential anti-bacterial
15 candidates. In one embodiment, the target sequence selected is specific to a single species, or even a single strain, such as, for example, the strain *B. fragilis* 14062. In a second embodiment, the target sequence is common to at least two species of bacteria. In a third embodiment, the target sequence is common to a family of bacteria. The target sequence may be a nucleic acid sequence or a polypeptide sequence. Methods employing
20 sequences common to more than one species of microorganism may be used to screen candidates for broad spectrum anti-bacterial activity.

The invention also provides methods for preventing or treating disease caused by certain bacteria, including *B. fragilis*, which are carried out by administering to an animal in need of such treatment, in particular a warm-blooded vertebrate, including but
25 not limited to birds and mammals, a compound that specifically inhibits or interferes with the function of a bacterial polypeptide or nucleic acid. In a particularly preferred embodiment, the mammal to be treated is human.

DETAILED DESCRIPTION OF THE INVENTION

The sequences of the present invention include the specific nucleic acid and amino acid sequences set forth in the Sequence Listing that forms a part of the present specification, and which are designated SEQ ID NO: 1 - SEQ ID NO: 10444. Use of the terms "SEQ ID NO: 1 - SEQ ID NO: 5222 ", " SEQ ID NO: 5223 - SEQ ID NO: 10444, "the sequences depicted in Table 2", etc., is intended, for convenience, to refer to each individual SEQ ID NO *individually*, and is not intended to refer to the genus of these sequences unless such reference would be indicated. In other words, it is a shorthand for listing all of these sequences individually. The invention encompasses each sequence individually, as well as any combination thereof.

DEFINITIONS

"Nucleic acid" or "polynucleotide" as used herein refers to purine- and pyrimidine-containing polymers of any length, either polyribonucleotides or polydeoxyribonucleotides or mixed polyribo-polydeoxyribo nucleotides. This includes single- and double-stranded molecules, i.e., DNA-DNA, DNA-RNA and RNA-RNA hybrids, as well as "protein nucleic acids" (PNA) formed by conjugating bases to an amino acid backbone. This also includes nucleic acids containing modified bases.

A nucleic acid or polypeptide sequence that is "derived from" a designated sequence refers to a sequence that corresponds to a region of the designated sequence. For nucleic acid sequences, this encompasses sequences that are homologous or complementary to the sequence, as well as "sequence-conservative variants" and "function-conservative variants." For polypeptide sequences, this encompasses "function-conservative variants." Sequence-conservative variants are those in which a change of one or more nucleotides in a given codon position results in no alteration in the amino acid encoded at that position. Function-conservative variants are those in which a given amino acid residue in a polypeptide has been changed without altering the overall conformation and function of the native polypeptide, including, but not limited to,

replacement of an amino acid with one having similar physico-chemical properties (such as, for example, acidic, basic, hydrophobic, and the like). "Function-conservative" variants also include any polypeptides that have the ability to elicit antibodies specific to a designated polypeptide.

5 An "*B. fragilis* -derived" nucleic acid or polypeptide sequence may or may not be present in other bacterial species, and may or may not be present in all *B. fragilis* strains. This term is intended to refer to the source from which the sequence was originally isolated. Thus, an *B. fragilis* -derived polypeptide, as used herein, may be used, e.g., as a target to screen for a broad spectrum antibacterial agent, to search for homologous

10 proteins in other species of bacteria or in eukaryotic organisms such as bacteria humans, etc.

A purified or isolated polypeptide or a substantially pure preparation of a polypeptide are used interchangeably herein and, as used herein, mean a polypeptide that has been separated from other proteins, lipids, and nucleic acids with which it naturally

15 occurs. Preferably, the polypeptide is also separated from substances, e.g., antibodies or gel matrix, e.g., polyacrylamide, which are used to purify it. Preferably, the polypeptide constitutes at least about 10, 20, 50 70, 80 or 95% dry weight of the purified preparation. Preferably, the preparation contains sufficient polypeptide to allow protein sequencing; at least about 1, 10, or preferably 100 mg of polypeptide.

20 A purified preparation of cells refers to, in the case of plant or animal cells, an *in vitro* preparation of cells and not an entire intact plant or animal. In the case of cultured cells or microbial cells, it consists of a preparation of at least about 10%, more preferably at least about 50%, of the subject cells.

A purified or isolated or a substantially pure nucleic acid, e.g., a substantially

25 pure DNA, (are terms used interchangeably herein) is a nucleic acid which is one or both of the following: not immediately contiguous with both of the coding sequences with which it is immediately contiguous (i.e., one at the 5' end and one at the 3' end) in the naturally-occurring genome of the organism from which the nucleic acid is derived; or

which is substantially free of a nucleic acid with which it occurs in the organism from which the nucleic acid is derived. The term includes, for example, a recombinant DNA which is incorporated into a vector, e.g., into an autonomously replicating plasmid or virus, or into the genomic DNA of a prokaryote or eukaryote, or which exists as a
5 separate molecule (e.g., a cDNA or a genomic DNA fragment produced by PCR or restriction endonuclease treatment) independent of other DNA sequences. Substantially pure DNA also includes a recombinant DNA which is part of a hybrid gene encoding additional *B. fragilis* DNA sequence.

A "contig" as used herein is a nucleic acid representing a continuous stretch of
10 genomic sequence of an organism.

An "open reading frame", also referred to herein as ORF, is a region of nucleic acid which encodes a polypeptide. This region may represent a portion of a coding sequence or a total sequence and can be determined from a stop to stop codon or from a start to stop codon.

15 As used herein, a "coding sequence" is a nucleic acid which is transcribed into messenger RNA and/or translated into a polypeptide when placed under the control of appropriate regulatory sequences. The boundaries of the coding sequence are determined by a translation start codon at the five prime terminus and a translation stop code at the three prime terminus. A coding sequence can include but is not limited to messenger
20 RNA, synthetic DNA, and recombinant nucleic acid sequences.

A "complement" of a nucleic acid as used herein refers to an anti-parallel or antisense sequence that participates in Watson-Crick base-pairing with the original sequence.

A "gene product" is a protein or structural RNA which is specifically encoded by
25 a gene.

As used herein, the term "probe" refers to a nucleic acid, peptide or other chemical entity which specifically binds to a molecule of interest. Probes are often associated with or capable of associating with a label. A label is a chemical moiety

capable of detection. Typical labels comprise dyes, radioisotopes, luminescent and chemiluminescent moieties, fluorophores, enzymes, precipitating agents, amplification sequences, and the like. Similarly, a nucleic acid, peptide or other chemical entity which specifically binds to a molecule of interest and immobilizes such molecule is referred
5 herein as a "capture ligand". Capture ligands are typically associated with or capable of associating with a support such as nitro-cellulose, glass, nylon membranes, beads, particles and the like. The specificity of hybridization is dependent on conditions such as the base pair composition of the nucleotides, and the temperature and salt concentration of the reaction. These conditions are readily discernable to one of ordinary skill in the art
10 using routine experimentation.

"Homologous" refers to the sequence similarity or sequence identity between two polypeptides or between two nucleic acid molecules. When a position in both of the two compared sequences is occupied by the same base or amino acid monomer subunit, e.g., if a position in each of two DNA molecules is occupied by adenine, then the molecules
15 are homologous at that position. The percent of homology between two sequences is a function of the number of matching or homologous positions shared by the two sequences divided by the number of positions compared x 100. For example, if 6 of 10 of the positions in two sequences are matched or homologous then the two sequences are 60% homologous. By way of example, the DNA sequences ATTGCC and TATGGC
20 share 50% homology. Generally, a comparison is made when two sequences are aligned to give maximum homology.

Nucleic acids are hybridizable to each other when at least one strand of a nucleic acid can anneal to the other nucleic acid under defined stringency conditions. Stringency of hybridization is determined by: (a) the temperature at which hybridization and/or
25 washing is performed; and (b) the ionic strength and polarity of the hybridization and washing solutions. Hybridization requires that the two nucleic acids contain complementary sequences; depending on the stringency of hybridization, however, mismatches may be tolerated. Typically, hybridization of two sequences at high

stringency (such as, for example, in a solution of 0.5X SSC, at 65° C) requires that the sequences be essentially completely homologous. Conditions of intermediate stringency (such as, for example, 2X SSC at 65 ° C) and low stringency (such as, for example 2X SSC at 55° C) require correspondingly less overall complementarity between the

5 hybridizing sequences. (1X SSC is 0.15 M NaCl, 0.015 M Na citrate).

The terms peptides, proteins, and polypeptides are used interchangeably herein.

As used herein, the term "surface protein" refers to all surface accessible proteins, e.g. inner and outer membrane proteins, proteins adhering to the cell wall, and secreted proteins.

10 A polypeptide has *B. fragilis* biological activity if it has one, two or preferably more of the following properties: (1) if when expressed in the course of an *B. fragilis* infection, it can promote, or mediate the attachment of *B. fragilis* to a cell; (2) it has an enzymatic activity, structural or regulatory function characteristic of an *B. fragilis* protein; (3) the gene which encodes it can rescue a lethal mutation in an *B. fragilis* gene.

15 A polypeptide has biological activity if it is an antagonist, agonist, or super-agonist of a polypeptide having one of the above-listed properties.

A biologically active fragment or analog is one having an *in vivo* or *in vitro* activity which is characteristic of the *B. fragilis* polypeptides of the invention contained in the Sequence Listing, or of other naturally occurring *B. fragilis* polypeptides, e.g., one

20 or more of the biological activities described herein. Especially preferred are fragments which exist *in vivo*, e.g., fragments which arise from post transcriptional processing or which arise from translation of alternatively spliced RNA's. Fragments include those expressed in native or endogenous cells as well as those made in expression systems, e.g., in CHO (Chinese Hamster Ovary) cells. Because peptides such as *B. fragilis*

25 polypeptides often exhibit a range of physiological properties and because such properties may be attributable to different portions of the molecule, a useful *B. fragilis* fragment or *B. fragilis* analog is one which exhibits a biological activity in any biological assay for *B. fragilis* activity. The fragment or analog possesses about 10%, preferably

about 40%, more preferably about 60%, 70%, 80% or 90% or greater of the activity of *B. fragilis*, in any *in vivo* or *in vitro* assay.

Analogues can differ from naturally occurring *B. fragilis* polypeptides in amino acid sequence or in ways that do not involve sequence, or both. Non-sequence modifications include changes in acetylation, methylation, phosphorylation, carboxylation, or glycosylation. Preferred analogues include *B. fragilis* polypeptides (or biologically active fragments thereof) whose sequences differ from the wild-type sequence by one or more conservative amino acid substitutions or by one or more non-conservative amino acid substitutions, deletions, or insertions which do not substantially diminish the biological activity of the *B. fragilis* polypeptide. Conservative substitutions typically include the substitution of one amino acid for another with similar characteristics, e.g., substitutions within the following groups: valine, glycine; glycine, alanine; valine, isoleucine, leucine; aspartic acid, glutamic acid; asparagine, glutamine; serine, threonine; lysine, arginine; and phenylalanine, tyrosine. Other conservative substitutions can be made in view of the table below.

TABLE 1

CONSERVATIVE AMINO ACID REPLACEMENTS

For Amino Acid	Code	Replace with any of
Alanine	A	D-Ala, Gly, beta-Ala, L-Cys, D-Cys
Arginine	R	D-Arg, Lys, D-Lys, homo-Arg, D-homo-Arg, Met, Ile, D-Met, D-Ile, Orn, D-Orn
Asparagine	N	D-Asn, Asp, D-Asp, Glu, D-Glu, Gln, D-Gln
Aspartic Acid	D	D-Asp, D-Asn, Asn, Glu, D-Glu, Gln, D-Gln
Cysteine	C	D-Cys, S-Me-Cys, Met, D-Met, Thr, D-Thr
Glutamine	Q	D-Gln, Asn, D-Asn, Glu, D-Glu, Asp, D-Asp
Glutamic Acid	E	D-Glu, D-Asp, Asp, Asn, D-Asn, Gln, D-Gln
Glycine	G	Ala, D-Ala, Pro, D-Pro, β -Ala, Acp
Isoleucine	I	D-Ile, Val, D-Val, Leu, D-Leu, Met, D-Met

Leucine	L	D-Leu, Val, D-Val, Leu, D-Leu, Met, D-Met
Lysine	K	D-Lys, Arg, D-Arg, homo-Arg, D-homo-Arg, Met, D-Met, Ile, D-Ile, Orn, D-Orn
Methionine	M	D-Met, S-Me-Cys, Ile, D-Ile, Leu, D-Leu, Val, D-Val
Phenylalanine	F	D-Phe, Tyr, D-Thr, L-Dopa, His, D-His, Trp, D-Trp, Trans-3,4, or 5-phenylproline, cis-3,4, or 5-phenylproline
Proline	P	D-Pro, L-I-thioazolidine-4-carboxylic acid, D-or L-1-oxazolidine-4-carboxylic acid
Serine	S	D-Ser, Thr, D-Thr, allo-Thr, Met, D-Met, Met(O), D-Met(O), L-Cys, D-Cys
Threonine	T	D-Thr, Ser, D-Ser, allo-Thr, Met, D-Met, Met(O), D-Met(O), Val, D-Val
Tyrosine	Y	D-Tyr, Phe, D-Phe, L-Dopa, His, D-His
Valine	V	D-Val, Leu, D-Leu, Ile, D-Ile, Met, D-Met

Other analogs within the invention are those with modifications which increase peptide stability; such analogs may contain, for example, one or more non-peptide bonds (which replace the peptide bonds) in the peptide sequence. Also included are: analogs

5 that include residues other than naturally occurring L-amino acids, e.g., D-amino acids or non-naturally occurring or synthetic amino acids, e.g., β or γ amino acids; and cyclic analogs.

As used herein, the term "fragment", as applied to an *B. fragilis* analog, will ordinarily be at least about 20 residues, more typically at least about 40 residues,

10 preferably at least about 60 residues in length. Fragments of *B. fragilis* polypeptides can be generated by methods known to those skilled in the art. The ability of an *Bacteroides* fragment to exhibit a biological activity of *B. fragilis* polypeptide can be assessed by methods known to those skilled in the art as described herein. Also included are *B. fragilis* polypeptides containing residues that are not required for biological activity of

15 the peptide or that result from alternative mRNA splicing or alternative protein processing events.

An "immunogenic component" as used herein is a moiety, such as an *B. fragilis* polypeptide, analog or fragment thereof, that is capable of eliciting a humoral and/or cellular immune response in a host animal.

An "antigenic component" as used herein is a moiety, such as an *B. fragilis* polypeptide, analog or fragment thereof, that is capable of binding to a specific antibody with sufficiently high affinity to form a detectable antigen-antibody complex.

The term "antibody" as used herein is intended to include fragments thereof which are specifically reactive with *B. fragilis* polypeptides.

As used herein, the term "cell-specific promoter" means a DNA sequence that serves as a promoter, i.e., regulates expression of a selected DNA sequence operably linked to the promoter, and which effects expression of the selected DNA sequence in specific cells of a tissue. The term also covers so-called "leaky" promoters, which regulate expression of a selected DNA primarily in one tissue, but cause expression in other tissues as well.

Misexpression, as used herein, refers to a non-wild type pattern of gene expression. It includes: expression at non-wild type levels, i.e., over or under expression; a pattern of expression that differs from wild type in terms of the time or stage at which the gene is expressed, e.g., increased or decreased expression (as compared with wild type) at a predetermined developmental period or stage; a pattern of expression that differs from wild type in terms of increased expression (as compared with wild type) in a predetermined cell type or tissue type; a pattern of expression that differs from wild type in terms of the splicing size, amino acid sequence, post-translational modification, or biological activity of the expressed polypeptide; a pattern of expression that differs from wild type in terms of the effect of an environmental stimulus or extracellular stimulus on expression of the gene, e.g., a pattern of increased or decreased expression (as compared with wild type) in the presence of an increase or decrease in the strength of the stimulus.

As used herein, "host cells" and other such terms denoting microorganisms or higher eukaryotic cell lines cultured as unicellular entities refers to cells which can

become or have been used as recipients for a recombinant vector or other transfer DNA, and include the progeny of the original cell which has been transfected. It is understood by individuals skilled in the art that the progeny of a single parental cell may not necessarily be completely identical in genomic or total DNA compliment to the original
5 parent, due to accident or deliberate mutation.

As used herein, the term "control sequence" refers to a nucleic acid having a base sequence which is recognized by the host organism to effect the expression of encoded sequences to which they are ligated. The nature of such control sequences differs depending upon the host organism; in prokaryotes, such control sequences generally
10 include a promoter, ribosomal binding site, terminators, and in some cases operators; in eukaryotes, generally such control sequences include promoters, terminators and in some instances, enhancers. The term control sequence is intended to include at a minimum, all components whose presence is necessary for expression, and may also include additional components whose presence is advantageous, for example, leader sequences.

15 As used herein, the term "operably linked" refers to sequences joined or ligated to function in their intended manner. For example, a control sequence is operably linked to coding sequence by ligation in such a way that expression of the coding sequence is achieved under conditions compatible with the control sequence and host cell.

The "metabolism" of a substance, as used herein, means any aspect of the
20 expression, function, action, or regulation of the substance. The metabolism of a substance includes modifications, e.g., covalent or non-covalent modifications of the substance. The metabolism of a substance includes modifications, e.g., covalent or non-covalent modification, the substance induces in other substances. The metabolism of a substance also includes changes in the distribution of the substance. The metabolism of a
25 substance includes changes the substance induces in the distribution of other substances.

A "sample" as used herein refers to a biological sample, such as, for example, tissue or fluid isolated from an individual (including without limitation plasma, serum,

cerebrospinal fluid, lymph, tears, saliva and tissue sections) or from *in vitro* cell culture constituents, as well as samples from the environment.

Technical and scientific terms used herein have the meanings commonly understood by one of ordinary skill in the art to which the present invention pertains, unless otherwise defined. Reference is made herein to various methodologies known to those of skill in the art. Publications and other materials setting forth such known methodologies to which reference is made are incorporated herein by reference in their entireties as though set forth in full. The practice of the invention will employ, unless otherwise indicated, conventional techniques of chemistry, molecular biology, microbiology, recombinant DNA, and immunology, which are within the skill of the art. Such techniques are explained fully in the literature. *See e.g.*, Sambrook, Fritsch, and Maniatis, *Molecular Cloning; Laboratory Manual* 2nd ed. (1989); *DNA Cloning*, Volumes I and II (D.N Glover ed. 1985); *Oligonucleotide Synthesis* (M.J. Gait ed, 1984); *Nucleic Acid Hybridization* (B.D. Hames & S.J. Higgins eds. 1984); the series, *Methods in Enzymology* (Academic Press, Inc.), particularly Vol. 154 and Vol. 155 (Wu and Grossman, eds.); *PCR-A Practical Approach* (McPherson, Quirke, and Taylor, eds., 1991); *Immunology*, 2d Edition, 1989, Roitt *et al.*, C.V. Mosby Company, and New York; *Advanced Immunology*, 2d Edition, 1991, Male *et al.*, Grower Medical Publishing, New York.; *DNA Cloning: A Practical Approach*, Volumes I and II, 1985 (D.N. Glover ed.); *Oligonucleotide Synthesis*, 1984, (M.L. Gait ed); *Transcription and Translation*, 1984 (Hames and Higgins eds.); *Animal Cell Culture*, 1986 (R.I. Freshney ed.); *Immobilized Cells and Enzymes*, 1986 (IRL Press); Perbal, 1984, *A Practical Guide to Molecular Cloning*; *Gene Transfer Vectors for Mammalian Cells*, 1987 (J. H. Miller and M. P. Calos eds., Cold Spring Harbor Laboratory); Martin J. Bishop, ed., *Guide to Human Genome Computing*, 2d Edition, Academic Press, San Diego, CA. (1998); and Leonard F. Peruski, Jr., and Anne Harwood Peruski, *The Internet and the New Biology: Tools for Genomic and Molecular Research*, American Society for Microbiology, Washington, D.C. (1997).

Any suitable materials and/or methods known to those of skill can be utilized in carrying out the present invention; however, preferred materials and/or methods are described. Materials, reagents and the like to which reference is made in the following description and examples are obtainable from commercial sources, unless otherwise
5 noted.

B. FRAGILIS GENOMIC SEQUENCE

This invention provides nucleotide sequences of the genome of *B. fragilis* which thus comprises a DNA sequence library of *B. fragilis* genomic DNA. The detailed
10 description that follows provides nucleotide sequences of *B. fragilis*, and also describes how the sequences were obtained and how ORFs and protein-coding sequences were identified. Also described are compositions and methods of using the disclosed *B. fragilis* sequences in methods including diagnostic and therapeutic applications. Furthermore, the library can be used as a database for identification and comparison of
15 medically important sequences in this and other strains of *B. fragilis*.

To determine the genomic sequence of *B. fragilis*, DNA from strain 14062 of *B. fragilis* was isolated after Zymolyase digestion, sodium dodecyl sulfate lysis, potassium acetate precipitation, phenol:chloroform extraction and ethanol precipitation (Soll, D.R., T. Srikantha and S.R. Lockhart: Characterizing Developmentally Regulated Genes in *B. fragilis*. In Microbial Genome Methods. K.W. Adolph, editor. CRC Press. New York.
20 p 17-37.). DNA was sheared hydrodynamically using an HPLC (Oefner, et. al., 1996) to an insert size of 2000-3000 bp. After size fractionation by gel electrophoresis the fragments were blunt-ended, ligated to adapter oligonucleotides and cloned into the pGTC (Thomann) vector to construct a "shotgun" subclone library.

25 DNA sequencing was achieved using established ABI sequencing methods on ABI377 automated DNA sequencers. The cloning and sequencing procedures are described in more detail in the Exemplification.

Individual sequence reads were assembled using PHRAP (P. Green, Abstracts of DOE Human Genome Program Contractor-Grantee Workshop V, Jan. 1996, p.157). The average contig length was about 3-4 kb.

All subsequent steps were based on sequencing by ABI377 automated DNA sequencing methods. The cloning and sequencing procedures are described in more detail in the Exemplification.

A variety of approaches may be used to order the contigs so as to obtain a continuous sequence representing the entire *B. fragilis* genome. Synthetic oligonucleotides are designed that are complementary to sequences at the end of each contig. These oligonucleotides may be hybridized to libraries of *B. fragilis* genomic DNA in, for example, lambda phage vectors or plasmid vectors to identify clones that contain sequences corresponding to the junctional regions between individual contigs. Such clones are then used to isolate template DNA and the same oligonucleotides are used as primers in polymerase chain reaction (PCR) to amplify junctional fragments, the nucleotide sequence of which is then determined.

The *B. fragilis* sequences were analyzed for the presence of open reading frames (ORFs) comprising at least 180 nucleotides. As a result of the analysis of ORFs based on stop-to-stop codon reads, it should be understood that these ORFs may not correspond to the ORF of a naturally-occurring *B. fragilis* polypeptide. These ORFs may contain start codons which indicate the initiation of protein synthesis of a naturally-occurring *B. fragilis* polypeptide. Such start codons within the ORFs provided herein were identified by those of ordinary skill in the relevant art, and the resulting ORF and the encoded *B. fragilis* polypeptide is within the scope of this invention. For example, within the ORFs a codon such as AUG or GUG (encoding methionine or valine) which is part of the initiation signal for protein synthesis were identified and the portion of an ORF to corresponding to a naturally-occurring *B. fragilis* polypeptide was recognized. The predicted coding regions were defined by evaluating the coding potential of such

sequences with the program GENEMARK™ (Borodovsky and McIninch, 1993, *Comp. .*
17:123).

Each predicted ORF amino acid sequence was compared with all sequences
found in current GENBANK, SWISS-PROT, and PIR databases using the BLAST
5 algorithm. BLAST identifies local alignments occurring by chance between the ORF
sequence and the sequence in the databank (Altschall et al., 1990, *L Mol. Biol.* 215:403-
410). Homologous ORFs (probabilities less than 10^{-5} by chance) and ORF's that are
probably non-homologous (probabilities greater than 10^{-5} by chance) but have good
codon usage were identified. Both homologous, sequences and non-homologous
10 sequences with good codon usage, are likely to encode proteins and are encompassed by
the invention.

B. FRAGILIS NUCLEIC ACIDS

The present invention provides a library of *B. fragilis* -derived nucleic acid
15 sequences. The libraries provide probes, primers, and markers which are used as markers
in epidemiological studies. The present invention also provides a library of *B. fragilis* -
derived nucleic acid sequences which comprise or encode targets for therapeutic drugs.

The nucleic acids of this invention may be obtained directly from the DNA of the
above referenced *B. fragilis* strain by using the polymerase chain reaction (PCR). See
20 "*PCR, A Practical Approach*" (McPherson, Quirke, and Taylor, eds., IRL Press, Oxford,
UK, 1991) for details about the PCR. High fidelity PCR is used to ensure a faithful DNA
copy prior to expression. In addition, the authenticity of amplified products is verified by
conventional sequencing methods. Clones carrying the desired sequences described in
this invention may also be obtained by screening the libraries by means of the PCR or by
25 hybridization of synthetic oligonucleotide probes to filter lifts of the library colonies or
plaques as known in the art (see, e.g., Sambrook et al., *Molecular Cloning, A Laboratory*
Manual 2nd edition, 1989, Cold Spring Harbor Press, NY).

It is also possible to obtain nucleic acids encoding *B. fragilis* polypeptides from a cDNA library in accordance with protocols herein described. A cDNA encoding an *B. fragilis* polypeptide can be obtained by isolating total mRNA from an appropriate strain. Double stranded cDNAs can then be prepared from the total mRNA. Subsequently, the cDNAs can be inserted into a suitable plasmid or viral (e.g., bacteriophage) vector using any one of a number of known techniques. Genes encoding *B. fragilis* polypeptides can also be cloned using established polymerase chain reaction techniques in accordance with the nucleotide sequence information provided by the invention. The nucleic acids of the invention can be DNA or RNA. Preferred nucleic acids of the invention are contained in the Sequence Listing.

The nucleic acids of the invention can also be chemically synthesized using standard techniques. Various methods of chemically synthesizing polydeoxynucleotides are known, including solid-phase synthesis which, like peptide synthesis, has been fully automated in commercially available DNA synthesizers (See e.g., Itakura et al. U.S. Patent No. 4,598,049; Caruthers et al. U.S. Patent No. 4,458,066; and Itakura U.S. Patent Nos. 4,401,796 and 4,373,071, incorporated by reference herein).

In another example, DNA can be chemically synthesized using, e.g., the phosphoramidite solid support method of Matteucci *et al.*, 1981, *J. Am. Chem. Soc.* 103:3185, the method of Yoo *et al.*, 1989, *J. Biol. Chem.* 264:17078, or other well known methods. This can be done by sequentially linking a series of oligonucleotide cassettes comprising pairs of synthetic oligonucleotides, as described below.

Nucleic acids isolated or synthesized in accordance with features of the present invention are useful, by way of example, without limitation, as probes, primers, capture ligands, antisense genes and for developing expression systems for the synthesis of proteins and peptides corresponding to such sequences. As probes, primers, capture ligands and antisense agents, the nucleic acid normally consists of all or part (approximately twenty or more nucleotides for specificity as well as the ability to form

stable hybridization products) of the nucleic acids of the invention contained in the Sequence Listing. These uses are described in further detail below.

PROBES

5 A nucleic acid isolated or synthesized in accordance with the sequence of the invention contained in the Sequence Listing can be used as a probe to specifically detect *B. fragilis*. With the sequence information set forth in the present application, sequences of twenty or more nucleotides are identified which provide the desired inclusivity and exclusivity with respect to *B. fragilis*, and extraneous nucleic acids likely to be
10 encountered during hybridization conditions. More preferably, the sequence will comprise at least about twenty to thirty nucleotides to convey stability to the hybridization product formed between the probe and the intended target molecules.

Sequences larger than 1000 nucleotides in length are difficult to synthesize but can be generated by recombinant DNA techniques. Individuals skilled in the art will
15 readily recognize that the nucleic acids, for use as probes, can be provided with a label to facilitate detection of a hybridization product.

Nucleic acid isolated and synthesized in accordance with the sequence of the invention contained in the Sequence Listing can also be useful as probes to detect homologous regions (especially homologous genes) of other *Bacteroides* species using
20 appropriate stringency hybridization conditions as described herein.

CAPTURE LIGAND

For use as a capture ligand, the nucleic acid selected in the manner described above with respect to probes, can be readily associated with a support. The manner in
25 which nucleic acid is associated with supports is well known. Nucleic acid having twenty or more nucleotides in a sequence of the invention contained in the Sequence Listing have utility to separate *B. fragilis* nucleic acid from one strain from the nucleic acid of other another strain as well as from other organisms. Nucleic acid having twenty

or more nucleotides in a sequence of the invention contained in the Sequence Listing can also have utility to separate other *Bacteroides* species from each other and from other organisms. Preferably, the sequence will comprise at least about twenty nucleotides to convey stability to the hybridization product formed between the probe and the intended target molecules. Sequences larger than 1000 nucleotides in length are difficult to synthesize but can be generated by recombinant DNA techniques.

PRIMERS

Nucleic acid isolated or synthesized in accordance with the sequences described herein have utility as primers for the amplification of *B. fragilis* nucleic acid. These nucleic acids may also have utility as primers for the amplification of nucleic acids in other *Bacteroides* species. With respect to polymerase chain reaction (PCR) techniques, nucleic acid sequences of ≥ 10 -15 nucleotides of the invention contained in the Sequence Listing have utility in conjunction with suitable enzymes and reagents to create copies of *B. fragilis* nucleic acid. More preferably, the sequence will comprise twenty or more nucleotides to convey stability to the hybridization product formed between the primer and the intended target molecules. Binding conditions of primers greater than 100 nucleotides are more difficult to control to obtain specificity. High fidelity PCR can be used to ensure a faithful DNA copy prior to expression. In addition, amplified products can be checked by conventional sequencing methods.

The copies can be used in diagnostic assays to detect specific sequences, including genes from *B. fragilis* and/or other *Bacteroides* species. The copies can also be incorporated into cloning and expression vectors to generate polypeptides corresponding to the nucleic acid synthesized by PCR, as is described in greater detail herein.

The nucleic acids of the present invention find use as templates for the recombinant production of *B. fragilis* -derived peptides or polypeptides

ANTISENSE

Nucleic acid or nucleic acid-hybridizing derivatives isolated or synthesized in accordance with the sequences described herein have utility as antisense agents to prevent the expression of *B. fragilis* genes. These sequences also have utility as
 5 antisense agents to prevent expression of genes of other *Bacteroides* species.

In one embodiment, nucleic acid or derivatives corresponding to *B. fragilis* nucleic acids is loaded into a suitable carrier such as a liposome or bacteriophage for introduction into bacterial cells. For example, a nucleic acid having twenty or more nucleotides is capable of binding to bacteria nucleic acid or bacteria messenger RNA.
 10 Preferably, the antisense nucleic acid is comprised of 20 or more nucleotides to provide necessary stability of a hybridization product of non-naturally occurring nucleic acid and bacterial nucleic acid and/or bacterial messenger RNA. Nucleic acid having a sequence greater than 1000 nucleotides in length is difficult to synthesize but can be generated by recombinant DNA techniques. Methods for loading antisense nucleic acid in liposomes
 15 is known in the art as exemplified by U.S. Patent 4,241,046 issued December 23, 1980 to Papahadjopoulos et al.

The present invention encompasses isolated polypeptides and nucleic acids derived from *B. fragilis* that are useful as reagents for diagnosis of bacterial infection, components of effective anti-bacterial vaccines, and/or as targets for anti-bacterial drugs,
 20 including anti-*B. fragilis* drugs.

EXPRESSION OF *B. FRAGILIS* NUCLEIC ACIDS

Table 2, which is appended herewith and which forms part of the present specification, provides a list of open reading frames (ORFs) in both strands and a
 25 putative identification of the particular function of a polypeptide which is encoded by each ORF, based on the homology match (determined by the BLASTP2 algorithm) of the predicted polypeptide with known proteins encoded by ORFs in other organisms. An ORF is a region of nucleic acid which encodes a polypeptide. This region may represent

a portion of a coding sequence or a total sequence and was determined from stop to stop codons. The first column contains a designation for the ORF ("ORF Name"). The second and third columns list the SEQ ID numbers for the nucleic acid ("NT ID") and amino acid ("AA ID") sequences corresponding to each ORF, respectively. The fourth and fifth

5 columns list the length of the nucleic acid ORF ("NT Length") and the length of the amino acid ORF ("AA Length"), respectively. The nucleotide sequence corresponding to each ORF begins at the first nucleotide immediately following a stop codon and ends at the nucleotide immediately preceding the next downstream stop codon in the same reading frame. It will be recognized by one skilled in the art that the natural translation

10 initiation sites will correspond to ATG, GTG, or TTG codons located within the ORFs. The natural initiation sites depend not only on the sequence of a start codon but also on the context of the DNA sequence adjacent to the start codon. Usually, a recognizable ribosome binding site is found within 20 nucleotides upstream from the initiation codon. In some cases where genes are translationally coupled and coordinately expressed

15 together in "operons", ribosome binding sites are not present, but the initiation codon of a downstream gene may occur very close to, or overlap, the stop codon of the an upstream gene in the same operon. The correct start codons can be generally identified without undue experimentation because only a few codons need be tested. It is recognized that the translational machinery in bacteria initiates all polypeptide chains with the amino

20 acid methionine, regardless of the sequence of the start codon. In some cases, polypeptides are post-translationally modified, resulting in an N-terminal amino acid other than methionine *in vivo*. The sixth and seventh columns provide metrics for assessing the likelihood of the homology match (determined by the BLASTP2 algorithm), as is known in the art, to the genes indicated in the description frame

25 ("Description") defined further below. These genes in the Description were identified when the designated ORF was compared against a comprehensive non-redundant protein database. Specifically, the sixth column represents the Blast Score ("Score") for the match (a higher score is a better match), and the seventh column represents the

probability ("Probability") for the match (the probability that such a match can have occurred by chance; the lower the value, the more likely the match is valid). If a BLASTP2 score of less than 100 was obtained, no value is reported in the table. The remaining fields below the columns contain additional information relating to the

5 potential function of the sequence based on the BLASTP2 analysis. Where a match was discovered, the field "Protein name" list the protein's name identified from the match. In addition, one skilled in the art would be able to identify the match and elucidate its function using the "Locus name" and where available the accession number, "Acc#" from the database. Lastly, one skilled in the art would appreciate the "Description" field to

10 further describe the potential function of the protein based on this analysis. This information allows one of ordinary skill in the art to determine a potential use for each identified coding sequence and, as a result, allows to use the polypeptides of the present invention for commercial and industrial purposes.

Using the information provided in SEQ ID NO: 1 - SEQ ID NO: 5222, SEQ ID

15 NO: 5223 - SEQ ID NO: 10444 and in Table 2 together with routine cloning and sequencing methods, one of ordinary skill in the art will be able to clone and sequence all the nucleic acid fragments of interest including open reading frames (ORFs) encoding a large variety of proteins of *B. fragilis*.

Nucleic acid isolated or synthesized in accordance with the sequences described

20 herein have utility to generate polypeptides. The nucleic acid of the invention exemplified in SEQ ID NO: 1 - SEQ ID NO: 5222 and in Table 2 or fragments of said nucleic acid encoding active portions of *B. fragilis* polypeptides can be cloned into suitable vectors or used to isolate nucleic acid. The isolated nucleic acid is combined with suitable DNA linkers and cloned into a suitable vector.

25 The function of a specific gene or operon can be ascertained by expression in a bacterial strain under conditions where the activity of the gene product(s) specified by the gene or operon in question can be specifically measured. Alternatively, a gene product may be produced in large quantities in an expressing strain for use as an antigen, an

industrial reagent, for structural studies, etc. This expression can be accomplished in a mutant strain which lacks the activity of the gene to be tested, or in a strain that does not produce the same gene product(s). This includes, but is not limited to, Eucaryotic species such as the yeast *Saccharomyces cerevisiae*, *Methanobacterium* strains or other Archaea, and Eubacteria such as *E. coli*, *B. Subtilis*, *S. Aureus*, *S. Pneumonia* or *Pseudomonas putida*. In some cases the expression host will utilize the natural *B. fragilis* promoter whereas in others, it will be necessary to drive the gene with a promoter sequence derived from the expressing organism (e.g., an *E. coli* beta-galactosidase promoter for expression in *E. coli*).

To express a gene product using the natural *B. fragilis* promoter, a procedure such as the following can be used. A restriction fragment containing the gene of interest, together with its associated natural promoter element and regulatory sequences (identified using the DNA sequence data) is cloned into an appropriate recombinant plasmid containing an origin of replication that functions in the host organism and an appropriate selectable marker. This can be accomplished by a number of procedures known to those skilled in the art. It is most preferably done by cutting the plasmid and the fragment to be cloned with the same restriction enzyme to produce compatible ends that can be ligated to join the two pieces together. The recombinant plasmid is introduced into the host organism by, for example, electroporation and cells containing the recombinant plasmid are identified by selection for the marker on the plasmid. Expression of the desired gene product is detected using an assay specific for that gene product.

In the case of a gene that requires a different promoter, the body of the gene (coding sequence) is specifically excised and cloned into an appropriate expression plasmid. This subcloning can be done by several methods, but is most easily accomplished by PCR amplification of a specific fragment and ligation into an expression plasmid after treating the PCR product with a restriction enzyme or exonuclease to create suitable ends for cloning.

A suitable host cell for expression of a gene can be any procaryotic or eucaryotic cell. Suitable methods for transforming host cells can be found in Sambrook et al. (Molecular Cloning: A Laboratory Manual, 2nd Edition, Cold Spring Harbor Laboratory Press (1989)), and other laboratory textbooks.

- 5 For example, a host cell transfected with a nucleic acid vector directing expression of a nucleotide sequence encoding an *B. fragilis* polypeptide can be cultured under appropriate conditions to allow expression of the polypeptide to occur. Suitable media for cell culture are well known in the art. Polypeptides of the invention can be isolated from cell culture medium, host cells, or both using techniques known in the art
- 10 for purifying proteins including ion-exchange chromatography, gel filtration chromatography, ultrafiltration, electrophoresis, and immunoaffinity purification with antibodies specific for such polypeptides. Additionally, in many situations, polypeptides can be produced by chemical cleavage of a native protein (e.g., tryptic digestion) and the cleavage products can then be purified by standard techniques.
- 15 In the case of membrane bound proteins, these can be isolated from a host cell by contacting a membrane-associated protein fraction with a detergent forming a solubilized complex, where the membrane-associated protein is no longer entirely embedded in the membrane fraction and is solubilized at least to an extent which allows it to be chromatographically isolated from the membrane fraction. Chromatographic techniques
- 20 which can be used in the final purification step are known in the art and include hydrophobic interaction, lectin affinity, ion exchange, dye affinity and immunoaffinity.

One strategy to maximize recombinant *B. fragilis* peptide expression in *E. coli* is to express the protein in a host bacteria with an impaired capacity to proteolytically cleave the recombinant protein (Gottesman, S., Gene Expression Technology: Methods

25 in Enzymology 185, Academic Press, San Diego, California (1990) 119-128). Another strategy would be to alter the nucleic acid encoding an *B. fragilis* peptide to be inserted into an expression vector so that the individual codons for each amino acid would be those preferentially utilized in highly expressed *E. coli* proteins (Wada et al., (1992) *Nuc.*

Acids Res. 20:2111-2118). Such alteration of nucleic acids of the invention can be carried out by standard DNA synthesis techniques.

The nucleic acids of the invention can also be chemically synthesized using standard techniques. Various methods of chemically synthesizing polydeoxynucleotides are known, including solid-phase synthesis which, like peptide synthesis, has been fully automated in commercially available DNA synthesizers (See, e.g., Itakura et al. U.S. Patent No. 4,598,049; Caruthers et al. U.S. Patent No. 4,458,066; and Itakura U.S. Patent Nos. 4,401,796 and 4,373,071, incorporated by reference herein).

The present invention provides a library of *B. fragilis* -derived nucleic acid sequences. The libraries provide probes, primers, and markers which can be used as markers in epidemiological studies. The present invention also provides a library of *B. fragilis* -derived nucleic acid sequences which comprise or encode targets for therapeutic drugs.

Nucleic acids comprising any of the sequences disclosed herein or sub-sequences thereof can be prepared by standard methods using the nucleic acid sequence information provided in SEQ ID NO: 1 - SEQ ID NO: 5222. For example, DNA can be chemically synthesized using, e.g., the phosphoramidite solid support method of Matteucci *et al.*, 1981, *J. Am. Chem. Soc.* 103:3185, the method of Yoo *et al.*, 1989, *J. Biol. Chem.* 264:17078, or other well known methods. This can be done by sequentially linking a series of oligonucleotide cassettes comprising pairs of synthetic oligonucleotides, as described below.

Of course, due to the degeneracy of the genetic code, many different nucleotide sequences can encode polypeptides having the amino acid sequences defined by SEQ ID NO: 5223 - SEQ ID NO: 10444 or sub-sequences thereof. The codons can be selected for optimal expression in prokaryotic or eukaryotic systems. Such degenerate variants are also encompassed by this invention.

Insertion of nucleic acids (typically DNAs) encoding the polypeptides of the invention into a vector is easily accomplished when the termini of both the DNAs and the

vector comprise compatible restriction sites. If this cannot be done, it may be necessary to modify the termini of the DNAs and/or vector by digesting back single-stranded DNA overhangs generated by restriction endonuclease cleavage to produce blunt ends, or to achieve the same result by filling in the single-stranded termini with an appropriate DNA
5 polymerase.

Alternatively, any site desired may be produced, e.g., by ligating nucleotide sequences (linkers) onto the termini. Such linkers may comprise specific oligonucleotide sequences that define desired restriction sites. Restriction sites can also be generated by the use of the polymerase chain reaction (PCR). See, e.g., Saiki *et al.*, 1988, *Science*
10 239:48. The cleaved vector and the DNA fragments may also be modified if required by homopolymeric tailing.

The nucleic acids of the invention may be isolated directly from cells. Alternatively, the polymerase chain reaction (PCR) method can be used to produce the nucleic acids of the invention, using either chemically synthesized strands or genomic
15 material as templates. Primers used for PCR can be synthesized using the sequence information provided herein and can further be designed to introduce appropriate new restriction sites, if desirable, to facilitate incorporation into a given vector for recombinant expression.

The nucleic acids of the present invention may be flanked by natural *B. fragilis*
20 regulatory sequences, or may be associated with heterologous sequences, including promoters, enhancers, response elements, signal sequences, polyadenylation sequences, introns, 5'- and 3'- noncoding regions, and the like. The nucleic acids may also be modified by many means known in the art. Non-limiting examples of such modifications include methylation, "caps", substitution of one or more of the naturally occurring
25 nucleotides with an analog, internucleotide modifications such as, for example, those with uncharged linkages (e.g., methyl phosphonates, phosphotriesters, phosphoroamidates, carbamates, etc.) and with charged linkages (e.g., phosphorothioates, phosphorodithioates, etc.). Nucleic acids may contain one or more additional covalently

linked moieties, such as, for example, proteins (e.g., nucleases, toxins, antibodies, signal peptides, poly-L-lysine, etc.), intercalators (e.g., acridine, psoralen, etc.), chelators (e.g., metals, radioactive metals, iron, oxidative metals, etc.), and alkylators. PNAs are also included. The nucleic acid may be derivatized by formation of a methyl or ethyl
5 phosphotriester or an alkyl phosphoramidate linkage. Furthermore, the nucleic acid sequences of the present invention may also be modified with a label capable of providing a detectable signal, either directly or indirectly. Exemplary labels include radioisotopes, fluorescent molecules, biotin, and the like.

The invention also provides nucleic acid vectors comprising the disclosed *B.*
10 *fragilis* -derived sequences or derivatives or fragments thereof. A large number of vectors, including plasmid and bacterial vectors, have been described for replication and/or expression in a variety of eukaryotic and prokaryotic hosts, and may be used for cloning or protein expression.

The encoded *B. fragilis* polypeptides may be expressed by using many known
15 vectors, such as pUC plasmids, pET plasmids (Novagen, Inc., Madison, WI), or pRSET or pREP (Invitrogen, San Diego, CA), and many appropriate host cells, using methods disclosed or cited herein or otherwise known to those skilled in the relevant art. The particular choice of vector/host is not critical to the practice of the invention.

Recombinant cloning vectors will often include one or more replication systems
20 for cloning or expression, one or more markers for selection in the host, e.g. antibiotic resistance, and one or more expression cassettes. The inserted *B. fragilis* coding sequences may be synthesized by standard methods, isolated from natural sources, or prepared as hybrids, etc. Ligation of the *B. fragilis* coding sequences to transcriptional regulatory elements and/or to other amino acid coding sequences may be achieved by
25 known methods. Suitable host cells may be transformed/transfected/infected as appropriate by any suitable method including electroporation, CaCl₂ mediated DNA uptake, bacterial infection, microinjection, microprojectile, or other established methods.

Appropriate host cells include bacteria, archebacteria, fungi, especially yeast, and plant and animal cells, especially mammalian cells. Of particular interest are *B. fragilis*, *E. coli*, *B. Subtilis*, *Saccharomyces cerevisiae*, *Saccharomyces carlsbergensis*, *Schizosaccharomyces pombe*, SF9 cells, C129 cells, 293 cells, *Neurospora*, and CHO
 5 cells, COS cells, HeLa cells, and immortalized mammalian myeloid and lymphoid cell lines. Preferred replication systems include M13, Cole1, SV40, baculovirus, lambda, adenovirus, and the like. A large number of transcription initiation and termination regulatory regions have been isolated and shown to be effective in the transcription and translation of heterologous proteins in the various hosts. Examples of these regions,
 10 methods of isolation, manner of manipulation, etc. are known in the art. Under appropriate expression conditions, host cells can be used as a source of recombinantly produced *B. fragilis* -derived peptides and polypeptides.

Advantageously, vectors may also include a transcription regulatory element (i.e., a promoter) operably linked to the *B. fragilis* portion. The promoter may optionally
 15 contain operator portions and/or ribosome binding sites. Non-limiting examples of bacterial promoters compatible with *E. coli* include: b-lactamase (penicillinase) promoter; lactose promoter; tryptophan (trp) promoter; araBAD (arabinose) operon promoter; lambda-derived P₁ promoter and N gene ribosome binding site; and the hybrid tac promoter derived from sequences of the trp and lac UV5 promoters. Non-limiting
 20 examples of yeast promoters include 3-phosphoglycerate kinase promoter, glyceraldehyde-3-phosphate dehydrogenase (GAPDH) promoter, galactokinase (GAL1) promoter, galactoeprimase promoter, and alcohol dehydrogenase (ADH) promoter. Suitable promoters for mammalian cells include without limitation viral promoters such as that from Simian Virus 40 (SV40), Rous sarcoma virus (RSV), adenovirus (ADV),
 25 and bovine papilloma virus (BPV). Mammalian cells may also require terminator sequences, polyA addition sequences and enhancer sequences to increase expression. Sequences which cause amplification of the gene may also be desirable. Furthermore, sequences that facilitate secretion of the recombinant product from cells, including, but

not limited to, bacteria, yeast, and animal cells, such as secretory signal sequences and/or prohormone pro region sequences, may also be included. These sequences are well described in the art.

Nucleic acids encoding wild-type or variant *B. fragilis* -derived polypeptides may also be introduced into cells by recombination events. For example, such a sequence can be introduced into a cell, and thereby effect homologous recombination at the site of an endogenous gene or a sequence with substantial identity to the gene. Other recombination-based methods such as nonhomologous recombinations or deletion of endogenous genes by homologous recombination may also be used.

The nucleic acids of the present invention find use as templates for the recombinant production of *B. fragilis* -derived peptides or polypeptides.

IDENTIFICATION AND USE OF *B. FRAGILIS* NUCLEIC ACID SEQUENCES

The disclosed *B. fragilis* polypeptide and nucleic acid sequences, or other sequences that are contained within ORFs, including complete protein-coding sequences, of which any of the disclosed *B. fragilis* -specific sequences forms a part, are useful as target components for diagnosis and/or treatment of *B. fragilis* - caused infection

It will be understood that the sequence of an entire protein-coding sequence of which each disclosed nucleic acid sequence forms a part can be isolated and identified based on each disclosed sequence. This can be achieved, for example, by using an isolated nucleic acid encoding the disclosed sequence, or fragments thereof, to prime a sequencing reaction with genomic *B. fragilis* DNA as template; this is followed by sequencing the amplified product. The isolated nucleic acid encoding the disclosed sequence, or fragments thereof, can also be hybridized to *B. fragilis* genomic libraries to identify clones containing additional complete segments of the protein-coding sequence of which the shorter sequence forms a part. Then, the entire protein-coding sequence, or fragments thereof, or nucleic acids encoding all or part of the sequence, or sequence-

conservative or function-conservative variants thereof, may be employed in practicing the present invention.

Preferred sequences are those that are useful in diagnostic and/or therapeutic applications. Diagnostic applications include without limitation nucleic-acid-based and antibody-based methods for detecting bacterial infection. Therapeutic applications include without limitation vaccines, passive immunotherapy, and drug treatments directed against gene products that are both unique to bacteria and essential for growth and/or replication of bacteria.

10 IDENTIFICATION OF NUCLEIC ACIDS ENCODING VACCINE COMPONENTS AND TARGETS FOR AGENTS EFFECTIVE AGAINST *B. FRAGILIS*

The disclosed *B. fragilis* genome sequence includes segments that direct the synthesis of ribonucleic acids and polypeptides, as well as origins of replication, promoters, other types of regulatory sequences, and intergenic nucleic acids. The invention encompasses nucleic acids encoding immunogenic components of vaccines and targets for agents effective against *B. fragilis*. Identification of said immunogenic components involved in the determination of the function of the disclosed sequences, which can be achieved using a variety of approaches. Non-limiting examples of these approaches are described briefly below.

20

HOMOLOGY TO KNOWN SEQUENCES:

Computer-assisted comparison of the disclosed *B. fragilis* sequences with previously reported sequences present in publicly available databases is useful for identifying functional *B. fragilis* nucleic acid and polypeptide sequences. It will be understood that protein-coding sequences, for example, may be compared as a whole, and that a high degree of sequence homology between two proteins (such as, for example, >80-90%) at the amino acid level indicates that the two proteins also possess some degree of functional homology, such as, for example, among enzymes involved in

metabolism, DNA synthesis, or cell wall synthesis, and proteins involved in transport, cell division, etc. In addition, many structural features of particular protein classes have been identified and correlate with specific consensus sequences, such as, for example, binding domains for nucleotides, DNA, metal ions, and other small molecules; sites for
5 covalent modifications such as phosphorylation, acylation, and the like; sites of protein:protein interactions, etc. These consensus sequences may be quite short and thus may represent only a fraction of the entire protein-coding sequence. Identification of such a feature in an *B. fragilis* sequence is therefore useful in determining the function of the encoded protein and identifying useful targets of antibacterial drugs.

10 Of particular relevance to the present invention are structural features that are common to secretory, transmembrane, and surface proteins, including secretion signal peptides and hydrophobic transmembrane domains. *B. fragilis* proteins identified as containing putative signal sequences and/or transmembrane domains are useful as immunogenic components of vaccines.

15 Targets for therapeutic drugs according to the invention include, but are not limited to, polypeptides of the invention, whether unique to *B. fragilis* or not, that are essential for growth and/or viability of *B. fragilis* under at least one growth condition. Polypeptides essential for growth and/or viability can be determined by examining the effect of deleting and/or disrupting the genes, i.e., by so-called gene "knockout".

20 Alternatively, genetic footprinting can be used (Smith *et al.*, 1995, *Proc. Natl. Acad. Sci. USA* 92:5479-6433; Published International Application WO 94/26933; U.S. Patent No. 5,612,180). Still other methods for assessing essentiality includes the ability to isolate conditional lethal mutations in the specific gene (e.g., temperature sensitive mutations). Other useful targets for therapeutic drugs, which include polypeptides that are not
25 essential for growth or viability *per se* but lead to loss of viability of the cell, can be used to target therapeutic agents to cells.

STRAIN-SPECIFIC SEQUENCES:

- Because of the evolutionary relationship between different *B. fragilis* strains, it is believed that the presently disclosed *B. fragilis* sequences are useful for identifying, and/or discriminating between, previously known and new *B. fragilis* strains. It is
- 5 believed that other *B. fragilis* strains will exhibit at least about 70% sequence homology with the presently disclosed sequence. Systematic and routine analyses of DNA sequences derived from samples containing *B. fragilis* strains, and comparison with the present sequence allows for the identification of sequences that can be used to discriminate between strains, as well as those that are common to all *B. fragilis* strains.
- 10 In one embodiment, the invention provides nucleic acids, including probes, and peptide and polypeptide sequences that discriminate between different strains of *B. fragilis*. Strain-specific components can also be identified functionally by their ability to elicit or react with antibodies that selectively recognize one or more *B. fragilis* strains.

- In another embodiment, the invention provides nucleic acids, including probes,
- 15 and peptide and polypeptide sequences that are common to all *B. fragilis* strains but are not found in other bacterial species.

B. FRAGILIS POLYPEPTIDES

- This invention encompasses isolated *B. fragilis* polypeptides encoded by the
- 20 disclosed *B. fragilis* genomic sequences, including the polypeptides of the invention contained in the Sequence Listing. Polypeptides of the invention are preferably at least about 5 amino acid residues in length. Using the DNA sequence information provided herein, the amino acid sequences of the polypeptides encompassed by the invention can be deduced using methods well-known in the art. It will be understood that the sequence
- 25 of an entire nucleic acid encoding an *B. fragilis* polypeptide can be isolated and identified based on an ORF that encodes only a fragment of the cognate protein-coding region. This can be achieved, for example, by using the isolated nucleic acid encoding

the ORF, or fragments thereof, to prime a polymerase chain reaction with genomic *B. fragilis* DNA as template; this is followed by sequencing the amplified product.

The polypeptides of the present invention, including function-conservative variants of the disclosed ORFs, may be isolated from wild-type or mutant *B. fragilis* cells, or from heterologous organisms or cells (including, but not limited to, bacteria, fungi, insect, plant, and mammalian cells) including *B. fragilis* into which an *B. fragilis* - derived protein-coding sequence has been introduced and expressed. Furthermore, the polypeptides may be part of recombinant fusion proteins.

B. fragilis polypeptides of the invention can be chemically synthesized using commercially automated procedures such as those referenced herein, including, without limitation, exclusive solid phase synthesis, partial solid phase methods, fragment condensation or classical solution synthesis. The polypeptides are preferably prepared by solid phase peptide synthesis as described by Merrifield, 1963, *J. Am. Chem. Soc.* 85:2149. The synthesis is carried out with amino acids that are protected at the alpha-amino terminus. Trifunctional amino acids with labile side-chains are also protected with suitable groups to prevent undesired chemical reactions from occurring during the assembly of the polypeptides. The alpha-amino protecting group is selectively removed to allow subsequent reaction to take place at the amino-terminus. The conditions for the removal of the alpha-amino protecting group do not remove the side-chain protecting groups.

Methods for polypeptide purification are well-known in the art, including, without limitation, preparative disc-gel electrophoresis, isoelectric focusing, HPLC, reversed-phase HPLC, gel filtration, ion exchange and partition chromatography, and countercurrent distribution. For some purposes, it is preferable to produce the polypeptide in a recombinant system in which the *B. fragilis* protein contains an additional sequence tag that facilitates purification, such as, but not limited to, a polyhistidine sequence. The polypeptide can then be purified from a crude lysate of the host cell by chromatography on an appropriate solid-phase matrix. Alternatively,

antibodies produced against an *B. fragilis* protein or against peptides derived therefrom can be used as purification reagents. Other purification methods are possible.

The present invention also encompasses derivatives and homologues of *B. fragilis* -encoded polypeptides. For some purposes, nucleic acid sequences encoding the peptides may be altered by substitutions, additions, or deletions that provide for functionally equivalent molecules, i.e., function-conservative variants. For example, one or more amino acid residues within the sequence can be substituted by another amino acid of similar properties, such as, for example, positively charged amino acids (arginine, lysine, and histidine); negatively charged amino acids (aspartate and glutamate); polar neutral amino acids; and non-polar amino acids.

The isolated polypeptides may be modified by, for example, phosphorylation, sulfation, acylation, or other protein modifications. They may also be modified with a label capable of providing a detectable signal, either directly or indirectly, including, but not limited to, radioisotopes and fluorescent compounds.

To identify *B. fragilis* -derived polypeptides for use in the present invention, essentially the complete genomic sequence of a virulent, methicillin-resistant isolate of *Bacteroides fragilis* isolate was analyzed. While, in very rare instances, a nucleic acid sequencing error may be revealed, resolving a rare sequencing error is well within the art, and such an occurrence will not prevent one skilled in the art from practicing the invention.

Also encompassed are any *B. fragilis* polypeptide sequences that are contained within the open reading frames (ORFs), including complete protein-coding sequences, of which any of SEQ ID NO: 1 - SEQ ID NO: 5222 forms a part. Table 2, which is appended herewith and which forms part of the present specification, provides a putative identification of the particular function of a polypeptide which is encoded by each ORF, based on the homology match (determined by the BLAST algorithm) of the predicted polypeptide with known proteins encoded by ORFs in other organisms. As a result, one

skilled in the art can use the polypeptides of the present invention for commercial and industrial purposes consistent with the type of putative identification of the polypeptide.

The present invention provides a library of *B. fragilis* -derived polypeptide sequences, and a corresponding library of nucleic acid sequences encoding the polypeptides, wherein the polypeptides themselves, or polypeptides contained within ORFs of which they form a part, comprise sequences that are contemplated for use as components of vaccines. Non-limiting examples of such sequences are listed by SEQ ID NO in Table 2, which is appended herewith and which forms part of the present specification.

10 The present invention also provides a library of *B. fragilis* -derived polypeptide sequences, and a corresponding library of nucleic acid sequences encoding the polypeptides, wherein the polypeptides themselves, or polypeptides contained within ORFs of which they form a part, comprise sequences lacking homology to any known prokaryotic or eukaryotic sequences. Such libraries provide probes, primers, and markers which can be used to diagnose *B. fragilis* infection, including use as markers in epidemiological studies. Non-limiting examples of such sequences are listed by SEQ ID NO in Table 2, which is appended hereto and part hereof.

The present invention also provides a library of *B. fragilis* -derived polypeptide sequences, and a corresponding library of nucleic acid sequences encoding the polypeptides, wherein the polypeptides themselves, or polypeptides contained within ORFs of which they form a part, comprise targets for therapeutic drugs.

SPECIFIC EXAMPLE: DETERMINATION OF *BACTEROIDES* PROTEIN ANTIGENS FOR ANTIBODY AND VACCINE DEVELOPMENT

25 The selection of *Bacteroides* protein antigens for vaccine development can be derived from the nucleic acids encoding *B. fragilis* polypeptides. First, the ORF's can be analyzed for homology to other known exported or membrane proteins and analyzed using the discriminant analysis described by Klein, et al. (Klein, P., Kanehsia, M., and

DeLisi, C. (1985) *Biochimica et Biophysica Acta* 815, 468-476) for predicting exported and membrane proteins.

Homology searches can be performed using the BLAST algorithm contained in the Wisconsin Sequence Analysis Package (Genetics Computer Group, University
 5 Research Park, 575 Science Drive, Madison, WI 53711) to compare each predicted ORF amino acid sequence with all sequences found in the current GenBank, SWISS-PROT and PIR databases. BLAST searches for local alignments between the ORF and the databank sequences and reports a probability score which indicates the probability of finding this sequence by chance in the database. ORF's with significant homology (e.g.
 10 probabilities lower than 1×10^{-6} that the homology is only due to random chance) to membrane or exported proteins represent protein antigens for vaccine development. Possible functions can be provided to *B. fragilis* genes based on sequence homology to genes cloned in other organisms.

Discriminant analysis (Klein, et al. supra) can be used to examine the ORF amino
 15 acid sequences. This algorithm uses the intrinsic information contained in the ORF amino acid sequence and compares it to information derived from the properties of known membrane and exported proteins. This comparison predicts which proteins will be exported, membrane associated or cytoplasmic. ORF amino acid sequences identified as exported or membrane associated by this algorithm are likely protein antigens for
 20 vaccine development.

PRODUCTION OF FRAGMENTS AND ANALOGS OF *B. FRAGILIS* NUCLEIC ACIDS AND POLYPEPTIDES

Based on the discovery of the *B. fragilis* gene products of the invention provided
 25 in the Sequence Listing, one skilled in the art can alter the disclosed structure of *B. fragilis* genes, e.g., by producing fragments or analogs, and test the newly produced structures for activity. Examples of techniques known to those skilled in the relevant art which allow the production and testing of fragments and analogs are discussed below.

These, or analogous methods can be used to make and screen libraries of polypeptides, e.g., libraries of random peptides or libraries of fragments or analogs of cellular proteins for the ability to bind *B. fragilis* polypeptides. Such screens are useful for the identification of inhibitors of *B. fragilis*.

5

GENERATION OF FRAGMENTS

Fragments of a protein can be produced in several ways, e.g., recombinantly, by proteolytic digestion, or by chemical synthesis. Internal or terminal fragments of a polypeptide can be generated by removing one or more nucleotides from one end (for a
10 terminal fragment) or both ends (for an internal fragment) of a nucleic acid which encodes the polypeptide. Expression of the mutagenized DNA produces polypeptide fragments. Digestion with "end-nibbling" endonucleases can thus generate DNAs which encode an array of fragments. DNAs which encode fragments of a protein can also be generated by random shearing, restriction digestion or a combination of the above-
15 discussed methods.

Fragments can also be chemically synthesized using techniques known in the art such as conventional Merrifield solid phase f-Moc or t-Boc chemistry. For example, peptides of the present invention may be arbitrarily divided into fragments of desired length with no overlap of the fragments, or divided into overlapping fragments of a
20 desired length.

ALTERATION OF NUCLEIC ACIDS AND POLYPEPTIDES: RANDOM METHODS

Amino acid sequence variants of a protein can be prepared by random mutagenesis of DNA which encodes a protein or a particular domain or region of a
25 protein. Useful methods include PCR mutagenesis and saturation mutagenesis. A library of random amino acid sequence variants can also be generated by the synthesis of a set of degenerate oligonucleotide sequences. (Methods for screening proteins in a library of variants are elsewhere herein).

PCR MUTAGENESIS

In PCR mutagenesis, reduced Taq polymerase fidelity is used to introduce random mutations into a cloned fragment of DNA (Leung et al., 1989, *Technique* 1:11-15). The DNA region to be mutagenized is amplified using the polymerase chain reaction (PCR) under conditions that reduce the fidelity of DNA synthesis by Taq DNA polymerase, e.g., by using a dGTP/dATP ratio of five and adding Mn^{2+} to the PCR reaction. The pool of amplified DNA fragments are inserted into appropriate cloning vectors to provide random mutant libraries.

10

SATURATION MUTAGENESIS

Saturation mutagenesis allows for the rapid introduction of a large number of single base substitutions into cloned DNA fragments (Mayers et al., 1985, *Science* 229:242). This technique includes generation of mutations, e.g., by chemical treatment or irradiation of single-stranded DNA *in vitro*, and synthesis of a complimentary DNA strand. The mutation frequency can be modulated by modulating the severity of the treatment, and essentially all possible base substitutions can be obtained. Because this procedure does not involve a genetic selection for mutant fragments both neutral substitutions, as well as those that alter function, are obtained. The distribution of point mutations is not biased toward conserved sequence elements.

20

DEGENERATE OLIGONUCLEOTIDES

A library of homologs can also be generated from a set of degenerate oligonucleotide sequences. Chemical synthesis of a degenerate sequences can be carried out in an automatic DNA synthesizer, and the synthetic genes then ligated into an appropriate expression vector. The synthesis of degenerate oligonucleotides is known in the art (see for example, Narang, SA (1983) *Tetrahedron* 39:3; Itakura et al. (1981) *Recombinant DNA, Proc 3rd Cleveland Sympos. Macromolecules*, ed. AG Walton,

25

Amsterdam: Elsevier pp273-289; Itakura et al. (1984) *Annu. Rev. Biochem.* 53:323; Itakura et al. (1984) *Science* 198:1056; Ike et al. (1983) *Nucleic Acid Res.* 11:477. Such techniques have been employed in the directed evolution of other proteins (see, for example, Scott et al. (1990) *Science* 249:386-390; Roberts et al. (1992) *PNAS* 89:2429-
 5 2433; Devlin et al. (1990) *Science* 249: 404-406; Cwirla et al. (1990) *PNAS* 87: 6378-6382; as well as U.S. Patents Nos. 5,223,409, 5,198,346, and 5,096,815).

ALTERATION OF NUCLEIC ACIDS AND POLYPEPTIDES: METHODS FOR DIRECTED MUTAGENESIS

10 Non-random or directed, mutagenesis techniques can be used to provide specific sequences or mutations in specific regions. These techniques can be used to create variants which include, e.g., deletions, insertions, or substitutions, of residues of the known amino acid sequence of a protein. The sites for mutation can be modified individually or in series, e.g., by (1) substituting first with conserved amino acids and
 15 then with more radical choices depending upon results achieved, (2) deleting the target residue, or (3) inserting residues of the same or a different class adjacent to the located site, or combinations of options 1-3.

ALANINE SCANNING MUTAGENESIS

20 Alanine scanning mutagenesis is a useful method for identification of certain residues or regions of the desired protein that are preferred locations or domains for mutagenesis, Cunningham and Wells (*Science* 244:1081-1085, 1989). In alanine scanning, a residue or group of target residues are identified (e.g., charged residues such as Arg, Asp, His, Lys, and Glu) and replaced by a neutral or negatively charged amino
 25 acid (most preferably alanine or polyalanine). Replacement of an amino acid can affect the interaction of the amino acids with the surrounding aqueous environment in or outside the cell. Those domains demonstrating functional sensitivity to the substitutions are then refined by introducing further or other variants at or for the sites of substitution.

Thus, while the site for introducing an amino acid sequence variation is predetermined, the nature of the mutation per se need not be predetermined. For example, to optimize the performance of a mutation at a given site, alanine scanning or random mutagenesis may be conducted at the target codon or region and the expressed desired protein subunit
5 variants are screened for the optimal combination of desired activity.

OLIGONUCLEOTIDE-MEDIATED MUTAGENESIS

Oligonucleotide-mediated mutagenesis is a useful method for preparing substitution, deletion, and insertion variants of DNA, see, e.g., Adelman et al., (*DNA*
10 2:183, 1983). Briefly, the desired DNA is altered by hybridizing an oligonucleotide encoding a mutation to a DNA template, where the template is the single-stranded form of a plasmid or bacteriophage containing the unaltered or native DNA sequence of the desired protein. After hybridization, a DNA polymerase is used to synthesize an entire second complementary strand of the template that will thus incorporate the
15 oligonucleotide primer, and will code for the selected alteration in the desired protein DNA. Generally, oligonucleotides of at least about 25 nucleotides in length are used. An optimal oligonucleotide will have 12 to 15 nucleotides that are completely complementary to the template on either side of the nucleotide(s) coding for the mutation. This ensures that the oligonucleotide will hybridize properly to the single-
20 stranded DNA template molecule. The oligonucleotides are readily synthesized using techniques known in the art such as that described by Crea et al. (*Proc. Natl. Acad. Sci. USA*, 75: 5765[1978]).

CASSETTE MUTAGENESIS

25 Another method for preparing variants, cassette mutagenesis, is based on the technique described by Wells et al. (*Gene*, 34:315[1985]). The starting material is a plasmid (or other vector) which includes the protein subunit DNA to be mutated. The codon(s) in the protein subunit DNA to be mutated are identified. There must be a

unique restriction endonuclease site on each side of the identified mutation site(s). If no such restriction sites exist, they may be generated using the above-described oligonucleotide-mediated mutagenesis method to introduce them at appropriate locations in the desired protein subunit DNA. After the restriction sites have been introduced into the plasmid, the plasmid is cut at these sites to linearize it. A double-stranded oligonucleotide encoding the sequence of the DNA between the restriction sites but containing the desired mutation(s) is synthesized using standard procedures. The two strands are synthesized separately and then hybridized together using standard techniques. This double-stranded oligonucleotide is referred to as the cassette. This cassette is designed to have 3' and 5' ends that are comparable with the ends of the linearized plasmid, such that it can be directly ligated to the plasmid. This plasmid now contains the mutated desired protein subunit DNA sequence.

COMBINATORIAL MUTAGENESIS

Combinatorial mutagenesis can also be used to generate mutants (Ladner et al., WO 88/06630). In this method, the amino acid sequences for a group of homologs or other related proteins are aligned, preferably to promote the highest homology possible. All of the amino acids which appear at a given position of the aligned sequences can be selected to create a degenerate set of combinatorial sequences. The variegated library of variants is generated by combinatorial mutagenesis at the nucleic acid level, and is encoded by a variegated gene library. For example, a mixture of synthetic oligonucleotides can be enzymatically ligated into gene sequences such that the degenerate set of potential sequences are expressible as individual peptides, or alternatively, as a set of larger fusion proteins containing the set of degenerate sequences.

OTHER MODIFICATIONS OF *B. FRAGILIS* NUCLEIC ACIDS AND POLYPEPTIDES

It is possible to modify the structure of an *B. fragilis* polypeptide for such purposes as increasing solubility, enhancing stability (e.g., shelf life *ex vivo* and
5 resistance to proteolytic degradation *in vivo*). A modified *B. fragilis* protein or peptide can be produced in which the amino acid sequence has been altered, such as by amino acid substitution, deletion, or addition as described herein.

An *B. fragilis* peptide can also be modified by substitution of cysteine residues preferably with alanine, serine, threonine, leucine or glutamic acid residues to minimize
10 dimerization via disulfide linkages. In addition, amino acid side chains of fragments of the protein of the invention can be chemically modified. Another modification is cyclization of the peptide.

In order to enhance stability and/or reactivity, an *B. fragilis* polypeptide can be modified to incorporate one or more polymorphisms in the amino acid sequence of the
15 protein resulting from any natural allelic variation. Additionally, D-amino acids, non-natural amino acids, or non-amino acid analogs can be substituted or added to produce a modified protein within the scope of this invention. Furthermore, an *B. fragilis* polypeptide can be modified using polyethylene glycol (PEG) according to the method of A. Schon and co-workers (Wie et al., *supra*) to produce a protein conjugated with PEG.
20 In addition, PEG can be added during chemical synthesis of the protein. Other modifications of *B. fragilis* proteins include reduction/alkylation (Tarr, *Methods of Protein Microcharacterization*, J. E. Silver ed., Humana Press, Clifton NJ 155-194 (1986)); acylation (Tarr, *supra*); chemical coupling to an appropriate carrier (Mishell and Shiigi, eds, *Selected Methods in Cellular Immunology*, WH Freeman, San Francisco, CA
25 (1980), U.S. Patent 4,939,239; or mild formalin treatment (Marsh, (1971) *Int. Arch. of Allergy and Appl. Immunol.*, 41: 199 - 215).

To facilitate purification and potentially increase solubility of an *B. fragilis* protein or peptide, it is possible to add an amino acid fusion moiety to the peptide

backbone. For example, hexa-histidine can be added to the protein for purification by immobilized metal ion affinity chromatography (Hochuli, E. et al., (1988) *Bio/Technology*, 6: 1321 - 1325). In addition, to facilitate isolation of peptides free of irrelevant sequences, specific endoprotease cleavage sites can be introduced between the

5 sequences of the fusion moiety and the peptide.

To potentially aid proper antigen processing of epitopes within an *B. fragilis* polypeptide, canonical protease sensitive sites can be engineered between regions, each comprising at least one epitope via recombinant or synthetic methods. For example, charged amino acid pairs, such as KK or RR, can be introduced between regions within a

10 protein or fragment during recombinant construction thereof. The resulting peptide can be rendered sensitive to cleavage by cathepsin and/or other trypsin-like enzymes which would generate portions of the protein containing one or more epitopes. In addition, such charged amino acid residues can result in an increase in the solubility of the peptide.

15 PRIMARY METHODS FOR SCREENING POLYPEPTIDES AND ANALOGS

Various techniques are known in the art for screening generated mutant gene products. Techniques for screening large gene libraries often include cloning the gene library into replicable expression vectors, transforming appropriate cells with the resulting library of vectors, and expressing the genes under conditions in which detection

20 of a desired activity, e.g., in this case, binding to *B. fragilis* polypeptide or an interacting protein, facilitates relatively easy isolation of the vector encoding the gene whose product was detected. Each of the techniques described below is amenable to high through-put analysis for screening large numbers of sequences created, e.g., by random mutagenesis techniques.

25

TWO HYBRID SYSTEMS

Two hybrid assays such as the system described below (as with the other screening methods described herein), can be used to identify polypeptides, e.g.,

fragments or analogs of a naturally-occurring *B. fragilis* polypeptide, e.g., of cellular proteins, or of randomly generated polypeptides which bind to an *B. fragilis* protein. (The *B. fragilis* domain is used as the bait protein and the library of variants are expressed as prey fusion proteins.) In an analogous fashion, a two hybrid assay (as with
 5 the other screening methods described herein), can be used to find polypeptides which bind an *B. fragilis* polypeptide.

DISPLAY LIBRARIES

In one approach to screening assays, the Bacteroides peptides are displayed on the
 10 surface of a cell or viral particle, and the ability of particular cells or viral particles to bind an appropriate receptor protein via the displayed product is detected in a "panning assay". For example, the gene library can be cloned into the gene for a surface membrane protein of a bacterial cell, and the resulting fusion protein detected by panning (Ladner et al., WO 88/06630; Fuchs et al. (1991) *Bio/Technology* 9:1370-1371; and
 15 Goward et al. (1992) *TIBS* 18:136-140). In a similar fashion, a detectably labeled ligand can be used to score for potentially functional peptide homologs. Fluorescently labeled ligands, e.g., receptors, can be used to detect homologs which retain ligand-binding activity. The use of fluorescently labeled ligands, allows cells to be visually inspected and separated under a fluorescence microscope, or, where the morphology of the cell
 20 permits, to be separated by a fluorescence-activated cell sorter.

A gene library can be expressed as a fusion protein on the surface of a viral particle. For instance, in the filamentous phage system, foreign peptide sequences can be expressed on the surface of infectious phage, thereby conferring two significant benefits. First, since these phage can be applied to affinity matrices at concentrations well over
 25 10^{13} phage per milliliter, a large number of phage can be screened at one time. Second, since each infectious phage displays a gene product on its surface, if a particular phage is recovered from an affinity matrix in low yield, the phage can be amplified by another round of infection. The group of almost identical *E. coli* filamentous phages, M13, fd.,

and fl, are most often used in phage display libraries. Either of the phage gIII or gVIII coat proteins can be used to generate fusion proteins without disrupting the ultimate packaging of the viral particle. Foreign epitopes can be expressed at the NH₂-terminal end of pIII and phage bearing such epitopes recovered from a large excess of phage
 5 lacking this epitope (Ladner et al. PCT publication WO 90/02909; Garrard et al., PCT publication WO 92/09690; Marks et al. (1992) *J. Biol. Chem.* 267:16007-16010; Griffiths et al. (1993) *EMBO J* 12:725-734; Clackson et al. (1991) *Nature* 352:624-628; and Barbas et al. (1992) *PNAS* 89:4457-4461).

A common approach uses the maltose receptor of *E. coli* (the outer membrane
 10 protein, LamB) as a peptide fusion partner (Charbit et al. (1986) *EMBO* 5, 3029-3037). Oligonucleotides have been inserted into plasmids encoding the LamB gene to produce peptides fused into one of the extracellular loops of the protein. These peptides are available for binding to ligands, e.g., to antibodies, and can elicit an immune response when the cells are administered to animals. Other cell surface proteins, e.g., OmpA
 15 (Schorr et al. (1991) *Vaccines* 9, pp. 387-392), PhoE (Agterberg, et al. (1990) *Gene* 88, 37-45), and PAL (Fuchs et al. (1991) *Bio/Tech* 9, 1369-1372), as well as large bacterial surface structures have served as vehicles for peptide display. Peptides can be fused to pilin, a protein which polymerizes to form the pilus-a conduit for interbacterial exchange of genetic information (Thiry et al. (1989) *Appl. Environ. Microbiol.* 55, 984-993).
 20 Because of its role in interacting with other cells, the pilus provides a useful support for the presentation of peptides to the extracellular environment. Another large surface structure used for peptide display is the bacterial motive organ, the flagellum. Fusion of peptides to the subunit protein flagellin offers a dense array of many peptide copies on the host cells (Kuwajima et al. (1988) *Bio/Tech.* 6, 1080-1083). Surface proteins of other
 25 bacterial species have also served as peptide fusion partners. Examples include the *Staphylococcus* protein A and the outer membrane IgA protease of *Neisseria* (Hansson et al. (1992) *J. Bacteriol.* 174, 4239-4245 and Klauser et al. (1990) *EMBO J.* 9, 1991-1999).

In the filamentous phage systems and the LamB system described above, the physical link between the peptide and its encoding DNA occurs by the containment of the DNA within a particle (cell or phage) that carries the peptide on its surface.

Capturing the peptide captures the particle and the DNA within. An alternative scheme
 5 uses the DNA-binding protein LacI to form a link between peptide and DNA (Cull *et al.* (1992) *PNAS USA* 89:1865-1869). This system uses a plasmid containing the LacI gene with an oligonucleotide cloning site at its 3'-end. Under the controlled induction by arabinose, a LacI-peptide fusion protein is produced. This fusion retains the natural ability of LacI to bind to a short DNA sequence known as LacO operator (LacO). By
 10 installing two copies of LacO on the expression plasmid, the LacI-peptide fusion binds tightly to the plasmid that encoded it. Because the plasmids in each cell contain only a single oligonucleotide sequence and each cell expresses only a single peptide sequence, the peptides become specifically and stably associated with the DNA sequence that directed its synthesis. The cells of the library are gently lysed and the peptide-DNA
 15 complexes are exposed to a matrix of immobilized receptor to recover the complexes containing active peptides. The associated plasmid DNA is then reintroduced into cells for amplification and DNA sequencing to determine the identity of the peptide ligands. As a demonstration of the practical utility of the method, a large random library of dodecapeptides was made and selected on a monoclonal antibody raised against the
 20 opioid peptide dynorphin B. A cohort of peptides was recovered, all related by a consensus sequence corresponding to a six-residue portion of dynorphin B. (Cull *et al.* (1992) *Proc. Natl. Acad. Sci. U.S.A.* 89:1869)

This scheme, sometimes referred to as peptides-on-plasmids, differs in two important ways from the phage display methods. First, the peptides are attached to the
 25 C-terminus of the fusion protein, resulting in the display of the library members as peptides having free carboxy termini. Both of the filamentous phage coat proteins, pIII and pVIII, are anchored to the phage through their C-termini, and the guest peptides are placed into the outward-extending N-terminal domains. In some designs, the phage-

displayed peptides are presented right at the amino terminus of the fusion protein.

(Cwirla, et al. (1990) *Proc. Natl. Acad. Sci. U.S.A.* 87, 6378-6382) A second difference is the set of biological biases affecting the population of peptides actually present in the libraries. The LacI fusion molecules are confined to the cytoplasm of the host cells. The phage coat fusions are exposed briefly to the cytoplasm during translation but are rapidly secreted through the inner membrane into the periplasmic compartment, remaining anchored in the membrane by their C-terminal hydrophobic domains, with the N-termini, containing the peptides, protruding into the periplasm while awaiting assembly into phage particles. The peptides in the LacI and phage libraries may differ significantly as a result of their exposure to different proteolytic activities. The phage coat proteins require transport across the inner membrane and signal peptidase processing as a prelude to incorporation into phage. Certain peptides exert a deleterious effect on these processes and are underrepresented in the libraries (Gallop et al. (1994) *J. Med. Chem.* 37(9):1233-1251). These particular biases are not a factor in the LacI display system.

The number of small peptides available in recombinant random libraries is enormous. Libraries of 10^7 - 10^9 independent clones are routinely prepared. Libraries as large as 10^{11} recombinants have been created, but this size approaches the practical limit for clone libraries. This limitation in library size occurs at the step of transforming the DNA containing randomized segments into the host bacterial cells. To circumvent this limitation, an *in vitro* system based on the display of nascent peptides in polysome complexes has recently been developed. This display library method has the potential of producing libraries 3-6 orders of magnitude larger than the currently available phage/phagemid or plasmid libraries. Furthermore, the construction of the libraries, expression of the peptides, and screening, is done in an entirely cell-free format.

In one application of this method (Gallop et al. (1994) *J. Med. Chem.* 37(9):1233-1251), a molecular DNA library encoding 10^{12} decapeptides was constructed and the library expressed in an *E. coli* S30 *in vitro* coupled transcription/translation system. Conditions were chosen to stall the ribosomes on the mRNA, causing the accumulation

of a substantial proportion of the RNA in polysomes and yielding complexes containing nascent peptides still linked to their encoding RNA. The polysomes are sufficiently robust to be affinity purified on immobilized receptors in much the same way as the more conventional recombinant peptide display libraries are screened. RNA from the bound
5 complexes is recovered, converted to cDNA, and amplified by PCR to produce a template for the next round of synthesis and screening. The polysome display method can be coupled to the phage display system. Following several rounds of screening, cDNA from the enriched pool of polysomes was cloned into a phagemid vector. This vector serves as both a peptide expression vector, displaying peptides fused to the coat
10 proteins, and as a DNA sequencing vector for peptide identification. By expressing the polysome-derived peptides on phage, one can either continue the affinity selection procedure in this format or assay the peptides on individual clones for binding activity in a phage ELISA, or for binding specificity in a completion phage ELISA (Barret, et al. (1992) *Anal. Biochem* 204,357-364). To identify the sequences of the active peptides
15 one sequences the DNA produced by the phagemid host.

SECONDARY SCREENING OF POLYPEPTIDES AND ANALOGS

The high through-put assays described above can be followed by secondary screens in order to identify further biological activities which will, e.g., allow one skilled
20 in the art to differentiate agonists from antagonists. The type of a secondary screen used will depend on the desired activity that needs to be tested. For example, an assay can be developed in which the ability to inhibit an interaction between a protein of interest and its respective ligand can be used to identify antagonists from a group of peptide fragments isolated though one of the primary screens described above.

25 Therefore, methods for generating fragments and analogs and testing them for activity are known in the art. Once the core sequence of interest is identified, it is routine for one skilled in the art to obtain analogs and fragments.

PEPTIDE MIMETICS OF *B. FRAGILIS* POLYPEPTIDES

The invention also provides for reduction of the protein binding domains of the subject *B. fragilis* polypeptides to generate mimetics, e.g. peptide or non-peptide agents. The peptide mimetics are able to disrupt binding of a polypeptide to its counter ligand, e.g., in the case of an *B. fragilis* polypeptide binding to a naturally occurring ligand. The critical residues of a subject *B. fragilis* polypeptide which are involved in molecular recognition of a polypeptide can be determined and used to generate *B. fragilis* -derived peptidomimetics which competitively or noncompetitively inhibit binding of the *B. fragilis* polypeptide with an interacting polypeptide (see, for example, European patent applications EP-412,762A and EP-B31,080A).

For example, scanning mutagenesis can be used to map the amino acid residues of a particular *B. fragilis* polypeptide involved in binding an interacting polypeptide, peptidomimetic compounds (e.g. diazepine or isoquinoline derivatives) can be generated which mimic those residues in binding to an interacting polypeptide, and which therefore can inhibit binding of an *B. fragilis* polypeptide to an interacting polypeptide and thereby interfere with the function of *B. fragilis* polypeptide. For instance, non-hydrolyzable peptide analogs of such residues can be generated using benzodiazepine (e.g., see Freidinger et al. in *Peptides: Chemistry and Biology*, G.R. Marshall ed., ESCOM Publisher: Leiden, Netherlands, 1988), azepine (e.g., see Huffman et al. in *Peptides: Chemistry and Biology*, G.R. Marshall ed., ESCOM Publisher: Leiden, Netherlands, 1988), substituted gamma lactam rings (Garvey et al. in *Peptides: Chemistry and Biology*, G.R. Marshall ed., ESCOM Publisher: Leiden, Netherlands, 1988), keto-methylene pseudopeptides (Ewenson et al. (1986) *J Med Chem* 29:295; and Ewenson et al. in *Peptides: Structure and Function* (Proceedings of the 9th American Peptide Symposium) Pierce Chemical Co. Rockland, IL, 1985), b-turn dipeptide cores (Nagai et al. (1985) *Tetrahedron Lett* 26:647; and Sato et al. (1986) *J Chem Soc Perkin Trans* 1:1231), and b-aminoalcohols (Gordon et al. (1985) *Biochem Biophys Res Commun* 126:419; and et al. (1986) *Biochem Biophys Res Commun* 134:71).

VACCINE FORMULATIONS FOR *B. FRAGILIS* NUCLEIC ACIDS AND POLYPEPTIDES

This invention also features vaccine compositions for protection against infection by *B. fragilis* or for treatment of *B. fragilis* infection. In one embodiment, the vaccine compositions contain one or more immunogenic components such as a surface protein from *B. fragilis*, or portion thereof, and a pharmaceutically acceptable carrier. Nucleic acids within the scope of the invention are exemplified by the nucleic acids of the invention contained in the Sequence Listing which encode *B. fragilis* surface proteins.

Any nucleic acid encoding an immunogenic *B. fragilis* protein, or portion thereof, which is capable of expression in a cell, can be used in the present invention. These vaccines have therapeutic and prophylactic utilities.

One aspect of the invention provides a vaccine composition for protection against infection by *B. fragilis* which contains at least one immunogenic fragment of an *B. fragilis* protein and a pharmaceutically acceptable carrier. Preferred fragments include peptides of at least about 10 amino acid residues in length, preferably about 10-20 amino acid residues in length, and more preferably about 12-16 amino acid residues in length.

Immunogenic components of the invention can be obtained, for example, by screening polypeptides recombinantly produced from the corresponding fragment of the nucleic acid encoding the full-length *B. fragilis* protein. In addition, fragments can be chemically synthesized using techniques known in the art such as conventional Merrifield solid phase f-Moc or t-Boc chemistry.

In one embodiment, immunogenic components are identified by the ability of the peptide to stimulate T cells. Peptides which stimulate T cells, as determined by, for example, T cell proliferation or cytokine secretion are defined herein as comprising at least one T cell epitope. T cell epitopes are believed to be involved in initiation and perpetuation of the immune response to the protein allergen which is responsible for the clinical symptoms of allergy. These T cell epitopes are thought to trigger early events at

the level of the T helper cell by binding to an appropriate HLA molecule on the surface of an antigen presenting cell, thereby stimulating the T cell subpopulation with the relevant T cell receptor for the epitope. These events lead to T cell proliferation, lymphokine secretion, local inflammatory reactions, recruitment of additional immune
 5 cells to the site of antigen/T cell interaction, and activation of the B cell cascade, leading to the production of antibodies. A T cell epitope is the basic element, or smallest unit of recognition by a T cell receptor, where the epitope comprises amino acids essential to receptor recognition (e.g., approximately 6 or 7 amino acid residues). Amino acid sequences which mimic those of the T cell epitopes are within the scope of this
 10 invention.

Screening immunogenic components can be accomplished using one or more of several different assays. For example, *in vitro*, peptide T cell stimulatory activity is assayed by contacting a peptide known or suspected of being immunogenic with an antigen presenting cell which presents appropriate MHC molecules in a T cell culture.
 15 Presentation of an immunogenic *B. fragilis* peptide in association with appropriate MHC molecules to T cells in conjunction with the necessary co-stimulation has the effect of transmitting a signal to the T cell that induces the production of increased levels of cytokines, particularly of interleukin-2 and interleukin-4. The culture supernatant can be obtained and assayed for interleukin-2 or other known cytokines. For example, any one
 20 of several conventional assays for interleukin-2 can be employed, such as the assay described in *Proc. Natl. Acad. Sci USA*, 86: 1333 (1989) the pertinent portions of which are incorporated herein by reference. A kit for an assay for the production of interferon is also available from Genzyme Corporation (Cambridge, MA).

Alternatively, a common assay for T cell proliferation entails measuring tritiated
 25 thymidine incorporation. The proliferation of T cells can be measured *in vitro* by determining the amount of ³H-labeled thymidine incorporated into the replicating DNA of cultured cells. Therefore, the rate of DNA synthesis and, in turn, the rate of cell division can be quantified.

Vaccine compositions of the invention containing immunogenic components (e.g., *B. fragilis* polypeptide or fragment thereof or nucleic acid encoding an *B. fragilis* polypeptide or fragment thereof) preferably include a pharmaceutically acceptable carrier. The term "pharmaceutically acceptable carrier" refers to a carrier that does not
5 cause an allergic reaction or other untoward effect in patients to whom it is administered. Suitable pharmaceutically acceptable carriers include, for example, one or more of water, saline, phosphate buffered saline, dextrose, glycerol, ethanol and the like, as well as combinations thereof. Pharmaceutically acceptable carriers may further comprise minor amounts of auxiliary substances such as wetting or emulsifying agents, preservatives or
10 buffers, which enhance the shelf life or effectiveness of the antibody. For vaccines of the invention containing *B. fragilis* polypeptides, the polypeptide is co-administered with a suitable adjuvant.

It will be apparent to those of skill in the art that the therapeutically effective amount of DNA or protein of this invention will depend, *inter alia*, upon the
15 administration schedule, the unit dose of antibody administered, whether the protein or DNA is administered in combination with other therapeutic agents, the immune status and health of the patient, and the therapeutic activity of the particular protein or DNA.

Vaccine compositions are conventionally administered parenterally, e.g., by injection, either subcutaneously or intramuscularly. Methods for intramuscular
20 immunization are described by Wolff et al. (1990) *Science* 247: 1465-1468 and by Sedegah et al. (1994) *Immunology* 91: 9866-9870. Other modes of administration include oral and pulmonary formulations, suppositories, and transdermal applications. Oral immunization is preferred over parenteral methods for inducing protection against infection by *B. fragilis*. Cain et. al. (1993) *Vaccine* 11: 637-642. Oral formulations
25 include such normally employed excipients as, for example, pharmaceutical grades of mannitol, lactose, starch, magnesium stearate, sodium saccharine, cellulose, magnesium carbonate, and the like.

The vaccine compositions of the invention can include an adjuvant, including, but not limited to aluminum hydroxide; N-acetyl-muramyl--L-threonyl-D-isoglutamine (thr-MDP); N-acetyl-nor-muramyl-L-alanyl-D-isoglutamine (CGP 11637, referred to as nor-MDP); N-acetylmuramyl-L-alanyl-D-isoglutaminyl-L-alanine-2-(1'-2'-dipalmitoyl-sn-glycero-3-hydroxyphosphoryloxy)-ethylamine (CGP 19835A, referred to as MTP-PE);

5 RIBI, which contains three components from bacteria; monophosphoryl lipid A; trehalose dimycolate; cell wall skeleton (MPL + TDM + CWS) in a 2% squalene/Tween 80 emulsion; and cholera toxin. Others which may be used are non-toxic derivatives of cholera toxin, including its B subunit, and/or conjugates or genetically engineered fusions

10 of the *B. fragilis* polypeptide with cholera toxin or its B subunit, procholeraenoid, fungal polysaccharides, including schizophyllan, muramyl dipeptide, muramyl dipeptide derivatives, phorbol esters, labile toxin of *E. coli*, non-*B. fragilis* bacterial lysates, block polymers or saponins.

Other suitable delivery methods include biodegradable microcapsules or immuno-

15 stimulating complexes (ISCOMs), cochleates, or liposomes, genetically engineered attenuated live vectors such as viruses or bacteria, and recombinant (chimeric) virus-like particles, e.g., bluetongue. The amount of adjuvant employed will depend on the type of adjuvant used. For example, when the mucosal adjuvant is cholera toxin, it is suitably used in an amount of 5 mg to 50 mg, for example 10 mg to 35 mg. When used in the

20 form of microcapsules, the amount used will depend on the amount employed in the matrix of the microcapsule to achieve the desired dosage. The determination of this amount is within the skill of a person of ordinary skill in the art.

Carrier systems in humans may include enteric release capsules protecting the antigen from the acidic environment of the stomach, and including *B. fragilis* polypeptide

25 in an insoluble form as fusion proteins. Suitable carriers for the vaccines of the invention are enteric coated capsules and polylactide-glycolide microspheres. Suitable diluents are 0.2 N NaHCO₃ and/or saline.

Vaccines of the invention can be administered as a primary prophylactic agent in adults or in children, as a secondary prevention, after successful eradication of *B. fragilis* in an infected host, or as a therapeutic agent in the aim to induce an immune response in a susceptible host to prevent infection by *B. fragilis*. The vaccines of the invention are administered in amounts readily determined by persons of ordinary skill in the art. Thus, for adults a suitable dosage will be in the range of 10 mg to 10 g, preferably 10 mg to 100 mg. A suitable dosage for adults will also be in the range of 5 mg to 500 mg. Similar dosage ranges will be applicable for children. Those skilled in the art will recognize that the optimal dose may be more or less depending upon the patient's body weight, disease, the route of administration, and other factors. Those skilled in the art will also recognize that appropriate dosage levels can be obtained based on results with known oral vaccines such as, for example, a vaccine based on an *E. coli* lysate (6 mg dose daily up to total of 540 mg) and with an enterotoxigenic *E. coli* purified antigen (4 doses of 1 mg) (Schulman et al., *J. Urol.* 150:917-921 (1993); Boedecker et al., *American Gastroenterological Assoc.* 999:A-222 (1993)). The number of doses will depend upon the disease, the formulation, and efficacy data from clinical trials. Without intending any limitation as to the course of treatment, the treatment can be administered over 3 to 8 doses for a primary immunization schedule over 1 month (Boedeker, *American Gastroenterological Assoc.* 888:A-222 (1993)).

In a preferred embodiment, a vaccine composition of the invention can be based on a killed whole *E. coli* preparation with an immunogenic fragment of an *B. fragilis* protein of the invention expressed on its surface or it can be based on an *E. coli* lysate, wherein the killed *E. coli* acts as a carrier or an adjuvant.

It will be apparent to those skilled in the art that some of the vaccine compositions of the invention are useful only for preventing *B. fragilis* infection, some are useful only for treating *B. fragilis* infection, and some are useful for both preventing and treating *B. fragilis* infection. In a preferred embodiment, the vaccine composition of the invention provides protection against *B. fragilis* infection by stimulating humoral

and/or cell-mediated immunity against *B. fragilis*. It should be understood that amelioration of any of the symptoms of *B. fragilis* infection is a desirable clinical goal, including a lessening of the dosage of medication used to treat *B. fragilis*-caused disease, or an increase in the production of antibodies in the serum or mucous of patients.

5

ANTIBODIES REACTIVE WITH *B. FRAGILIS* POLYPEPTIDES

The invention also includes antibodies specifically reactive with the subject *B. fragilis* polypeptide. Anti-protein/anti-peptide antisera or monoclonal antibodies can be made by standard protocols (See, for example, *Antibodies: A Laboratory Manual* ed. by Harlow and Lane (Cold Spring Harbor Press: 1988)). A mammal such as a mouse, a hamster or rabbit can be immunized with an immunogenic form of the peptide. Techniques for conferring immunogenicity on a protein or peptide include conjugation to carriers or other techniques well known in the art. An immunogenic portion of the subject *B. fragilis* polypeptide can be administered in the presence of adjuvant. The progress of immunization can be monitored by detection of antibody titers in plasma or serum. Standard ELISA or other immunoassays can be used with the immunogen as antigen to assess the levels of antibodies.

In a preferred embodiment, the subject antibodies are immunospecific for antigenic determinants of the *B. fragilis* polypeptides of the invention, e.g. antigenic determinants of a polypeptide of the invention contained in the Sequence Listing, or a closely related human or non-human mammalian homolog (e.g., 90% homologous, more preferably at least about 95% homologous). In yet a further preferred embodiment of the invention, the anti-*B. fragilis* antibodies do not substantially cross react (i.e., react specifically) with a protein which is for example, less than 80% percent homologous to a sequence of the invention contained in the Sequence Listing. By "not substantially cross react", it is meant that the antibody has a binding affinity for a non-homologous protein which is less than 10 percent, more preferably less than 5 percent, and even more preferably less than 1 percent, of the binding affinity for a protein of the invention

contained in the Sequence Listing. In a most preferred embodiment, there is no cross-reactivity between bacterial and mammalian antigens.

The term antibody as used herein is intended to include fragments thereof which are also specifically reactive with *B. fragilis* polypeptides. Antibodies can be fragmented using conventional techniques and the fragments screened for utility in the same manner as described above for whole antibodies. For example, F(ab')₂ fragments can be generated by treating antibody with pepsin. The resulting F(ab')₂ fragment can be treated to reduce disulfide bridges to produce Fab' fragments. The antibody of the invention is further intended to include bispecific and chimeric molecules having an anti-*B. fragilis* portion.

Both monoclonal and polyclonal antibodies (Ab) directed against *B. fragilis* polypeptides or *B. fragilis* polypeptide variants, and antibody fragments such as Fab' and F(ab')₂, can be used to block the action of *B. fragilis* polypeptide and allow the study of the role of a particular *B. fragilis* polypeptide of the invention in aberrant or unwanted intracellular signaling, as well as the normal cellular function of the *B. fragilis* and by microinjection of anti-*B. fragilis* polypeptide antibodies of the present invention.

Antibodies which specifically bind *B. fragilis* epitopes can also be used in immunohistochemical staining of tissue samples in order to evaluate the abundance and pattern of expression of *B. fragilis* antigens. Anti-*B. fragilis* polypeptide antibodies can be used diagnostically in immuno-precipitation and immuno-blotting to detect and evaluate *B. fragilis* levels in tissue or bodily fluid as part of a clinical testing procedure. Likewise, the ability to monitor *B. fragilis* polypeptide levels in an individual can allow determination of the efficacy of a given treatment regimen for an individual afflicted with such a disorder. The level of an *B. fragilis* polypeptide can be measured in cells found in bodily fluid, such as in urine samples or can be measured in tissue, such as produced by gastric biopsy. Diagnostic assays using anti-*B. fragilis* antibodies can include, for example, immunoassays designed to aid in early diagnosis of *B. fragilis* infections. The

present invention can also be used as a method of detecting antibodies contained in samples from individuals infected by this bacterium using specific *B. fragilis* antigens.

Another application of anti-*B. fragilis* polypeptide antibodies of the invention is in the immunological screening of cDNA libraries constructed in expression vectors such as λ gt11, λ gt18-23, λ ZAP, and λ ORF8. Messenger libraries of this type, having coding sequences inserted in the correct reading frame and orientation, can produce fusion proteins. For instance, λ gt11 will produce fusion proteins whose amino termini consist of β -galactosidase amino acid sequences and whose carboxy termini consist of a foreign polypeptide. Antigenic epitopes of a subject *B. fragilis* polypeptide can then be detected with antibodies, as, for example, reacting nitrocellulose filters lifted from infected plates with anti-*B. fragilis* polypeptide antibodies. Phage, scored by this assay, can then be isolated from the infected plate. Thus, the presence of *B. fragilis* gene homologs can be detected and cloned from other species, and alternate isoforms (including splicing variants) can be detected and cloned.

KITS CONTAINING NUCLEIC ACIDS, POLYPEPTIDES OR ANTIBODIES OF THE INVENTION

The nucleic acid, polypeptides and antibodies of the invention can be combined with other reagents and articles to form kits. Kits for diagnostic purposes typically comprise the nucleic acid, polypeptides or antibodies in vials or other suitable vessels. Kits typically comprise other reagents for performing hybridization reactions, polymerase chain reactions (PCR), or for reconstitution of lyophilized components, such as aqueous media, salts, buffers, and the like. Kits may also comprise reagents for sample processing such as detergents, chaotropic salts and the like. Kits may also comprise immobilization means such as particles, supports, wells, dipsticks and the like. Kits may also comprise labeling means such as dyes, developing reagents, radioisotopes, fluorescent agents, luminescent or chemiluminescent agents, enzymes, intercalating agents and the like. With the nucleic acid and amino acid sequence information provided

herein, individuals skilled in art can readily assemble kits to serve their particular purpose. Kits further can include instructions for use.

BIO CHIP TECHNOLOGY

5 The nucleic acid sequence of the present invention may be used to detect *B. fragilis* or other species of *Bacteroides* acid sequence using bio chip technology. Bio chips containing arrays of nucleic acid sequence can also be used to measure expression of genes of *B. fragilis* or other species of *Bacteroides*. For example, to diagnose a patient with a *B. fragilis* or other *Bacteroides* infection, a sample from a human or animal can be
10 used as a probe on a bio chip containing an array of nucleic acid sequence from the present invention. In addition, a sample from a disease state can be compared to a sample from a non-disease state which would help identify a gene that is up-regulated or expressed in the disease state. This would provide valuable insight as to the mechanism by which the disease manifests. Changes in gene expression can also be used to identify
15 critical pathways involved in drug transport or metabolism, and may enable the identification of novel targets involved in virulence or host cell interactions involved in maintenance of an infection. Procedures using such techniques have been described by Brown *et al.*, 1995, *Science* 270: 467-470.

 Bio chips can also be used to monitor the genetic changes of potential therapeutic
20 compounds including, deletions, insertions or mismatches. Once the therapeutic is added to the patient, changes to the genetic sequence can be evaluated for its efficacy. In addition, the nucleic acid sequence of the present invention can be used to determine essential genes in cell cycling. As described in Iyer *et al.*, 1999 (*Science*, 283:83-87) genes essential in the cell cycle can be identified using bio chips. Furthermore, the
25 present invention provides nucleic acid sequence which can be used with bio chip technology to understand regulatory networks in bacteria, measure the response to environmental signals or drugs as in drug screening, and study virulence induction.

(Mons *et al.*, 1998, *Nature Biotechnology*, 16: 45-48. Patents teaching this technology include U.S. Patents 5445934, 5744305, and 5800992.

DRUG SCREENING ASSAYS USING *B. FRAGILIS* POLYPEPTIDES

5 By making available purified and recombinant *B. fragilis* polypeptides, the present invention provides assays which can be used to screen for drugs which are either agonists or antagonists of the normal cellular function, in this case, of the subject *B. fragilis* polypeptides, or of their role in intracellular signaling. Such inhibitors or potentiators may be useful as new therapeutic agents to combat *B. fragilis* infections in
10 humans. A variety of assay formats will suffice and, in light of the present inventions, will be comprehended by the person skilled in the art.

In many drug screening programs which test libraries of compounds and natural extracts, high throughput assays are desirable in order to maximize the number of compounds surveyed in a given period of time. Assays which are performed in cell-free
15 systems, such as may be derived with purified or semi-purified proteins, are often preferred as "primary" screens in that they can be generated to permit rapid development and relatively easy detection of an alteration in a molecular target which is mediated by a test compound. Moreover, the effects of cellular toxicity and/or bioavailability of the test compound can be generally ignored in the *in vitro* system, the assay instead being
20 focused primarily on the effect of the drug on the molecular target as may be manifest in an alteration of binding affinity with other proteins or change in enzymatic properties of the molecular target. Accordingly, in an exemplary screening assay of the present invention, the compound of interest is contacted with an isolated and purified *B. fragilis* polypeptide.

25 Screening assays can be constructed *in vitro* with a purified *B. fragilis* polypeptide or fragment thereof, such as an *B. fragilis* polypeptide having enzymatic activity, such that the activity of the polypeptide produces a detectable reaction product. The efficacy of the compound can be assessed by generating dose response curves from

data obtained using various concentrations of the test compound. Moreover, a control assay can also be performed to provide a baseline for comparison. Suitable products include those with distinctive absorption, fluorescence, or chemi-luminescence properties, for example, because detection may be easily automated. A variety of synthetic or naturally occurring compounds can be tested in the assay to identify those which inhibit or potentiate the activity of the *B. fragilis* polypeptide. Some of these active compounds may directly, or with chemical alterations to promote membrane permeability or solubility, also inhibit or potentiate the same activity (e.g., enzymatic activity) in whole, live *B. fragilis* cells.

10

OVEREXPRESSION ASSAYS

Overexpression assays are based on the premise that overproduction of a protein would lead to a higher level of resistance to compounds that selectively interfere with the function of that protein. Overexpression assays may be used to identify compounds that interfere with the function of virtually any type of protein, including without limitation enzymes, receptors, DNA- or RNA-binding proteins, or any proteins that are directly or indirectly involved in regulating cell growth.

Typically, two bacterial strains are constructed. One contains a single copy of the gene of interest, and a second contains several copies of the same gene. Identification of useful inhibitory compounds of this type of assay is based on a comparison of the activity of a test compound in inhibiting growth and/or viability of the two strains. The method involves constructing a nucleic acid vector that directs high level expression of a particular target nucleic acid. The vectors are then transformed into host cells in single or multiple copies to produce strains that express low to moderate and high levels of protein encoding by the target sequence (strain A and B, respectively). Nucleic acid comprising sequences encoding the target gene can, of course, be directly integrated into the host cell.

Large numbers of compounds (or crude substances which may contain active compounds) are screened for their effect on the growth of the two strains. Agents which interfere with an unrelated target equally inhibit the growth of both strains. Agents which interfere with the function of the target at high concentration should inhibit the growth of both strains. It should be possible, however, to titrate out the inhibitory effect of the compound in the overexpressing strain. That is, if the compound is affecting the particular target that is being tested, it should be possible to inhibit the growth of strain A at a concentration of the compound that allows strain B to grow.

Alternatively, a bacterial strain is constructed that contains the gene of interest under the control of an inducible promoter. Identification of useful inhibitory agents using this type of assay is based on a comparison of the activity of a test compound in inhibiting growth and/or viability of this strain under both inducing and non-inducing conditions. The method involves constructing a nucleic acid vector that directs high-level expression of a particular target nucleic acid. The vector is then transformed into host cells that are grown under both non-inducing and inducing conditions (conditions A and B, respectively).

Large numbers of compounds (or crude substances which may contain active compounds) are screened for their effect on growth under these two conditions. Agents that interfere with the function of the target should inhibit growth under both conditions. It should be possible, however, to titrate out the inhibitory effect of the compound in the overexpressing strain. That is, if the compound is affecting the particular target that is being tested, it should be possible to inhibit growth under condition A at a concentration that allows the strain to grow under condition B.

LIGAND-BINDING ASSAYS

Many of the targets according to the invention have functions that have not yet been identified. Ligand-binding assays are useful to identify inhibitor compounds that interfere with the function of a particular target, even when that function is unknown.

These assays are designed to detect binding of test compounds to particular targets. The detection may involve direct measurement of binding. Alternatively, indirect indications of binding may involve stabilization of protein structure or disruption of a biological function. Non-limiting examples of useful ligand-binding assays are detailed below.

5 A useful method for the detection and isolation of binding proteins is the Biomolecular Interaction Assay (BIAcore) system developed by Pharmacia Biosensor and described in the manufacturer's protocol (LKB Pharmacia, Sweden). The BIAcore system uses an affinity purified anti-GST antibody to immobilize GST-fusion proteins onto a sensor chip. The sensor utilizes surface plasmon resonance which is an optical
10 phenomenon that detects changes in refractive indices. In accordance with the practice of the invention, a protein of interest is coated onto a chip and test compounds are passed over the chip. Binding is detected by a change in the refractive index (surface plasmon resonance).

 A different type of ligand-binding assay involves scintillation proximity assays
15 (SPA, described in U.S. Patent No. 4,568,649).

 Another type of ligand binding assay, also undergoing development, is based on the fact that proteins containing mitochondrial targeting signals are imported into isolated mitochondria *in vitro* (Hurt *et al.*, 1985, *Embo J.* 4:2061-2068; Eilers and Schatz, *Nature*, 1986, 322:228-231). In a mitochondrial import assay, expression vectors are constructed
20 in which nucleic acids encoding particular target proteins are inserted downstream of sequences encoding mitochondrial import signals. The chimeric proteins are synthesized and tested for their ability to be imported into isolated mitochondria in the absence and presence of test compounds. A test compound that binds to the target protein should inhibit its uptake into isolated mitochondria *in vitro*.

25 Another ligand-binding assay is the yeast two-hybrid system (Fields and Song, 1989, *Nature* 340:245-246). The yeast two-hybrid system takes advantage of the properties of the GAL4 protein of the yeast *Saccharomyces cerevisiae*. The GAL4 protein is a transcriptional activator required for the expression of genes encoding

enzymes of galactose utilization. This protein consists of two separable and functionally essential domains: an N-terminal domain which binds to specific DNA sequences (UAS_G); and a C-terminal domain containing acidic regions, which is necessary to activate transcription. The native GAL4 protein, containing both domains, is a potent activator of transcription when yeast are grown on galactose media. The N-terminal domain binds to DNA in a sequence-specific manner but is unable to activate transcription. The C-terminal domain contains the activating regions but cannot activate transcription because it fails to be localized to UAS_G. In the two-hybrid system, a system of two hybrid proteins containing parts of GAL4: (1) a GAL4 DNA-binding domain fused to a protein 'X' and (2) a GAL4 activation region fused to a protein 'Y'. If X and Y can form a protein-protein complex and reconstitute proximity of the GAL4 domains, transcription of a gene regulated by UAS_G occurs. Creation of two hybrid proteins, each containing one of the interacting proteins X and Y, allows the activation region of UAS_G to be brought to its normal site of action.

The binding assay described in Fodor *et al.*, 1991, *Science* 251:767-773, which involves testing the binding affinity of test compounds for a plurality of defined polymers synthesized on a solid substrate, may also be useful.

Compounds which bind to the polypeptides of the invention are potentially useful as antibacterial agents for use in therapeutic compositions.

Pharmaceutical formulations suitable for antibacterial therapy comprise the antibacterial agent in conjunction with one or more biologically acceptable carriers. Suitable biologically acceptable carriers include, but are not limited to, phosphate-buffered saline, saline, deionized water, or the like. Preferred biologically acceptable carriers are physiologically or pharmaceutically acceptable carriers.

The antibacterial compositions include an antibacterial effective amount of active agent. Antibacterial effective amounts are those quantities of the antibacterial agents of the present invention that afford prophylactic protection against bacterial infections or which result in amelioration or cure of an existing bacterial infection. This antibacterial

effective amount will depend upon the agent, the location and nature of the infection, and the particular host. The amount can be determined by experimentation known in the art, such as by establishing a matrix of dosages and frequencies and comparing a group of experimental units or subjects to each point in the matrix.

5 The antibacterial active agents or compositions can be formed into dosage unit forms, such as for example, creams, ointments, lotions, powders, liquids, tablets, capsules, suppositories, sprays, aerosols or the like. If the antibacterial composition is formulated into a dosage unit form, the dosage unit form may contain an antibacterial effective amount of active agent. Alternatively, the dosage unit form may include less
10 than such an amount if multiple dosage unit forms or multiple dosages are to be used to administer a total dosage of the active agent. Dosage unit forms can include, in addition, one or more excipient(s), diluent(s), disintegrant(s), lubricant(s), plasticizer(s), colorant(s), dosage vehicle(s), absorption enhancer(s), stabilizer(s), bactericide(s), or the like.

15 For general information concerning formulations, see, e.g., Gilman et al. (eds.), 1990, *Goodman and Gilman's: The Pharmacological Basis of Therapeutics*, 8th ed., Pergamon Press; and *Remington's Pharmaceutical Sciences*, 17th ed., 1990, Mack Publishing Co., Easton, PA; Avis et al. (eds.), 1993, *Pharmaceutical Dosage Forms: Parenteral Medications*, Dekker, New York; Lieberman et al (eds.), 1990,
20 *Pharmaceutical Dosage Forms: Disperse Systems*, Dekker, New York.

 The antibacterial agents and compositions of the present invention are useful for preventing or treating *B. fragilis* infections. Infection prevention methods incorporate a prophylactically effective amount of an antibacterial agent or composition. A prophylactically effective amount is an amount effective to prevent *B. fragilis* infection
25 and will depend upon the specific bacterial strain, the agent, and the host. These amounts can be determined experimentally by methods known in the art and as described above.

B. fragilis infection treatment methods incorporate a therapeutically effective amount of an antibacterial agent or composition. A therapeutically effective amount is an amount sufficient to ameliorate or eliminate the infection. The prophylactically and/or therapeutically effective amounts can be administered in one administration or over
5 repeated administrations. Therapeutic administration can be followed by prophylactic administration, once the initial bacterial infection has been resolved.

The antibacterial agents and compositions can be administered topically or systemically. Topical application is typically achieved by administration of creams, ointments, lotions, or sprays as described above. Systemic administration includes both
10 oral and parental routes. Parental routes include, without limitation, subcutaneous, intramuscular, intraperitoneal, intravenous, transdermal, inhalation and intranasal administration.

EXEMPLIFICATION

15

CLONING AND SEQUENCING *B. FRAGILIS* GENOMIC SEQUENCE

This invention provides nucleotide sequences of the genome of *B. fragilis* which thus comprises a DNA sequence library of *B. fragilis* genomic DNA. The detailed description that follows provides nucleotide sequences of *B. fragilis*, and also describes
20 how the sequences were obtained and how ORFs (Open Reading Frames) and protein-coding sequences can be identified. Also described are methods of using the disclosed *B. fragilis* sequences in methods including diagnostic and therapeutic applications. Furthermore, the library can be used as a database for identification and comparison of medically important sequences in this and other strains of *B. fragilis* as well as other
25 species of *Bacteroides*.

Chromosomal DNA from strain 14062 of *B. fragilis* was isolated after Zymolyase digestion, sodium dodecyl sulfate lysis, potassium acetate precipitation, phenol:chloroform extraction and ethanol precipitation (Soll, D.R., T. Srikantha and S.R.

Lockhart: Characterizing Developmentally Regulated Genes in *B. fragilis*. In Microbial Genome Methods. K.W. Adolph, editor. CRC Press. New York. p 17-37.). Genomic *B. fragilis* DNA was hydrodynamically sheared in an HPLC and then separated on a standard 1% agarose gel. Fractions corresponding to 2500-3000 bp in length were
 5 excised from the gel and purified by the GeneClean procedure (Bio101, Inc.).

The purified DNA fragments were then blunt-ended using T4 DNA polymerase. The healed DNA was then ligated to unique *Bst*XI-linker adapters (5'-GTCTTCACCACGGGG-3' and 5'-GTGGTGAAGAC-3' in 100-1000 fold molar excess). These linkers are complimentary to the *Bst*XI-cut pGTC vector, while the
 10 overhang is not self-complimentary. Therefore, the linkers will not concatamerize nor will the cut-vector religate itself easily. The linker-adapted inserts were separated from the unincorporated linkers on a 1% agarose gel and purified using GeneClean. The linker-adapted inserts were then ligated to *Bst*XI-cut vector to construct a "shotgun" subclone libraries.

15 Only major modifications to the protocols are highlighted. Briefly, the library was then transformed into DH5 α competent cells (Gibco/BRL, DH5 α transformation protocol). It was assessed by plating onto antibiotic plates containing ampicillin and IPTG/Xgal. The plates were incubated overnight at 37°C. Transformants were then used for plating of clones and picking for sequencing. The cultures were grown overnight at
 20 37°C. DNA was purified using a silica bead DNA preparation (Engelstein, 1996) method. In this manner, 25 μ g of DNA was obtained per clone.

These purified DNA samples were then sequenced using primarily ABI dye-terminator chemistry. All subsequent steps were based on sequencing by ABI377 automated DNA sequencing methods. The ABI dye terminator sequence reads were run
 25 on ABI377 machines and the data was transferred to UNIX machines following lane tracking of the gels. Base calls and quality scores were determined using the program PHRED (Ewing et al., 1998, Genome Res. 8: 175-185; Ewing and Green, 1998, Genome Res. 8: 685-734). Reads were assembled using PHRAP (P. Green, Abstracts of DOE

Human Genome Program Contractor-Grantee Workshop V, Jan. 1996, p.157) with default program parameters and quality scores. The initial assembly was done at 7.8 fold coverage and yielded 223 contigs.

Finishing can follow the initial assembly. Missing mates (sequences from clones
5 that only gave reads from one end of the *Bacteroides* DNA inserted in the plasmid) can be identified and sequenced with ABI technology to allow the identification of additional overlapping contigs.

End-sequencing of randomly picked genomic lambda was also performed. Sequencing on a both sides was done for all lambda sequences. The lambda library
10 backbone helped to verify the integrity of the assembly and allowed closure of some of the physical gaps. Primers for walking off the ends of contigs would be selected using pick_primer (a GTC program) near the ends of the clones to facilitate gap closure. These walks can be sequenced using the selected clones and primers. These data are then reassembled with PHRAP. Additional sequencing using PCR-generated templates and
15 screened and/or unscreened lambda templates can be done in addition.

To identify *B. fragilis* polypeptides the complete genomic sequence of *B. fragilis* were analyzed essentially as follows: First, all possible stop-to- stop open reading frames (ORFs) greater than 180 nucleotides in all six reading frames were translated into amino acid sequences. Second, the identified ORFs were analyzed for homology to known
20 (archeabacter, prokaryotic and eukaryotic) protein sequences. Third, the coding potential of non-homologous sequences were evaluated with the program GENEMARK™ (Borodovsky and McIninch, 1993, Comp. Chem. 17:123).

IDENTIFICATION, CLONING AND EXPRESSION OF *B. FRAGILIS* NUCLEIC 25 ACIDS

Expression and purification of the *B. fragilis* polypeptides of the invention can be performed essentially as outlined below.

To facilitate the cloning, expression and purification of membrane and secreted proteins from *B. fragilis*, a gene expression system, such as the pET System (Novagen), for cloning and expression of recombinant proteins in *E. coli*, is selected. Also, a DNA sequence encoding a peptide tag, the His-Tag, is fused to the 3' end of DNA sequences of interest in order to facilitate purification of the recombinant protein products. The 3' end is selected for fusion in order to avoid alteration of any 5' terminal signal sequence.

PCR AMPLIFICATION AND CLONING OF NUCLEIC ACIDS CONTAINING ORF'S ENCODING ENZYMES

10 Nucleic acids chosen (for example, from the nucleic acids set forth in SEQ ID NO: 1 - SEQ ID NO: 5222 for cloning from the 14062 strain of *B. fragilis* are prepared for amplification cloning by polymerase chain reaction (PCR). Synthetic oligonucleotide primers specific for the 5' and 3' ends of open reading frames (ORFs) are designed and purchased from GibcoBRL Life Technologies (Gaithersburg, MD, USA). All forward primers (specific for the 5' end of the sequence) are designed to include an NcoI cloning site at the extreme 5' terminus. These primers are designed to permit initiation of protein translation at a methionine residue followed by a valine residue and the coding sequence for the remainder of the native *B. fragilis* DNA sequence. All reverse primers (specific for the 3' end of any *B. fragilis* ORF) include a EcoRI site at the extreme 5' terminus to permit cloning of each *B. fragilis* sequence into the reading frame of the pET-28b. The pET-28b vector provides sequence encoding an additional 20 carboxy-terminal amino acids including six histidine residues (at the extreme C-terminus), which comprise the His-Tag.

25 Genomic DNA prepared from the 14062 strain of *B. fragilis* is used as the source of template DNA for PCR amplification reactions (Current Protocols in Molecular Biology, John Wiley and Sons, Inc., F. Ausubel et al., eds., 1994). To amplify a DNA sequence containing an *B. fragilis* ORF, genomic DNA (50 nanograms) is introduced into a reaction vial containing 2 mM MgCl₂, 1 micromolar synthetic oligonucleotide

primers (forward and reverse primers) complementary to and flanking a defined *B. fragilis* ORF, 0.2 mM of each deoxynucleotide triphosphate; dATP, dGTP, dCTP, dTTP and 2.5 units of heat stable DNA polymerase (Amplitaq, Roche Molecular Systems, Inc., Branchburg, NJ, USA) in a final volume of 100 microliters.

5 Upon completion of thermal cycling reactions, each sample of amplified DNA is washed and purified using the Qiaquick Spin PCR purification kit (Qiagen, Gaithersburg, MD, USA). All amplified DNA samples are subjected to digestion with the restriction endonucleases, e.g., NcoI and EcoRI (New England BioLabs, Beverly, MA, USA)(Current Protocols in Molecular Biology, John Wiley and Sons, Inc., F. Ausubel et al., eds., 1994). DNA samples are then subjected to electrophoresis on 1.0 % NuSeive
10 (FMC BioProducts, Rockland, ME USA) agarose gels. DNA is visualized by exposure to ethidium bromide and long wave uv irradiation. DNA contained in slices isolated from the agarose gel is purified using the Bio 101 GeneClean Kit protocol (Bio 101 Vista, CA, USA).

15

CLONING OF *B. FRAGILIS* NUCLEIC ACIDS INTO AN EXPRESSION VECTOR

The pET-28b vector is prepared for cloning by digestion with restriction endonucleases, e.g., NcoI and EcoRI (Current Protocols in Molecular Biology, John Wiley and Sons, Inc., F. Ausubel et al., eds., 1994). The pET-28a vector, which encodes
20 a His-Tag that can be fused to the 5' end of an inserted gene, is prepared by digestion with appropriate restriction endonucleases.

Following digestion, DNA inserts are cloned (Current Protocols in Molecular Biology, John Wiley and Sons, Inc., F. Ausubel et al., eds., 1994) into the previously digested pET-28b expression vector. Products of the ligation reaction are then used to
25 transform the BL21 strain of *E. coli* (Current Protocols in Molecular Biology, John Wiley and Sons, Inc., F. Ausubel et al., eds., 1994) as described below.

TRANSFORMATION OF COMPETENT BACTERIA WITH RECOMBINANT PLASMIDS

Competent bacteria, *E. coli* strain BL21 or *E. coli* strain BL21(DE3), are transformed with recombinant pET expression plasmids carrying the cloned *B. fragilis* sequences according to standard methods (Current Protocols in Molecular, John Wiley and Sons, Inc., F. Ausubel et al., eds., 1994). Briefly, 1 microliter of ligation reaction is mixed with 50 microliters of electrocompetent cells and subjected to a high voltage pulse, after which, samples are incubated in 0.45 milliliters SOC medium (0.5% yeast extract, 2.0 % tryptone, 10 mM NaCl, 2.5 mM KCl, 10 mM MgCl₂, 10 mM MgSO₄ and 20, mM glucose) at 37°C with shaking for 1 hour. Samples are then spread on LB agar plates containing 25 microgram/ml kanamycin sulfate for growth overnight. Transformed colonies of BL21 are then picked and analyzed to evaluate cloned inserts as described below.

15 IDENTIFICATION OF RECOMBINANT EXPRESSION VECTORS WITH *B. FRAGILIS* NUCLEIC ACIDS

Individual BL21 clones transformed with recombinant pET-28b *B. fragilis* ORFs are analyzed by PCR amplification of the cloned inserts using the same forward and reverse primers, specific for each *B. fragilis* sequence, that were used in the original PCR amplification cloning reactions. Successful amplification verifies the integration of the *B. fragilis* sequences in the expression vector (Current Protocols in Molecular Biology, John Wiley and Sons, Inc., F. Ausubel et al., eds., 1994).

25 ISOLATION AND PREPARATION OF NUCLEIC ACIDS FROM TRANSFORMANTS

Individual clones of recombinant pET-28b vectors carrying properly cloned *B. fragilis* ORFs are picked and incubated in 5 mls of LB broth plus 25 microgram/ml

kanamycin sulfate overnight. The following day plasmid DNA is isolated and purified using the Qiagen plasmid purification protocol (Qiagen Inc., Chatsworth, CA, USA).

EXPRESSION OF RECOMBINANT *B. FRAGILIS* SEQUENCES IN *E. COLI*

- 5 The pET vector can be propagated in any *E. coli* K-12 strain e.g. HMS174, HB101, JM109, DH5, etc. for the purpose of cloning or plasmid preparation. Hosts for expression include *E. coli* strains containing a chromosomal copy of the gene for T7 RNA polymerase. These hosts are lysogens of bacteriophage DE3, a lambda derivative that carries the *lacI* gene, the *lacUV5* promoter and the gene for T7 RNA polymerase. T7
- 10 RNA polymerase is induced by addition of isopropyl-B-D-thiogalactoside (IPTG), and the T7 RNA polymerase transcribes any target plasmid, such as pET-28b, carrying its gene of interest. Strains used include: BL21(DE3) (Studier, F.W., Rosenberg, A.H., Dunn, J.J., and Dubendorff, J.W. (1990) *Meth. Enzymol.* 185, 60-89).

- To express recombinant *B. fragilis* sequences, 50 nanograms of plasmid DNA
- 15 isolated as described above is used to transform competent BL21(DE3) bacteria as described above (provided by Novagen as part of the pET expression system kit). The *lacZ* gene (beta-galactosidase) is expressed in the pET-System as described for the *B. fragilis* recombinant constructions. Transformed cells are cultured in SOC medium for 1 hour, and the culture is then plated on LB plates containing 25 micrograms/ml
- 20 kanamycin sulfate. The following day, bacterial colonies are pooled and grown in LB medium containing kanamycin sulfate (25 micrograms/ml) to an optical density at 600 nM of 0.5 to 1.0 O.D. units, at which point, 1 millimolar IPTG was added to the culture for 3 hours to induce gene expression of the *B. fragilis* recombinant DNA constructions.

- After induction of gene expression with IPTG, bacteria are pelleted by
- 25 centrifugation in a Sorvall RC-3B centrifuge at 3500 x g for 15 minutes at 4°C. Pellets are resuspended in 50 milliliters of cold 10 mM Tris-HCl, pH 8.0, 0.1 M NaCl and 0.1 mM EDTA (STE buffer). Cells are then centrifuged at 2000 x g for 20 min at 4°C. Wet pellets are weighed and frozen at -80°C until ready for protein purification.

A variety of methodologies known in the art can be utilized to purify the isolated proteins. (Current Protocols in Protein Science, John Wiley and Sons, Inc., J. E. Coligan et al., eds., 1995). For example, the frozen cells may be thawed, resuspended in buffer and ruptured by several passages through a small volume microfluidizer (Model M-110S,
 5 Microfluidics International Corporation, Newton, MA). The resultant homogenate may be centrifuged to yield a clear supernatant (crude extract) and following filtration the crude extract may be fractionated over columns. Fractions may be monitored by absorbance at OD₂₈₀ nm. and peak fractions may analyzed by SDS-PAGE

The concentrations of purified protein preparations may be quantified
 10 spectrophotometrically using absorbance coefficients calculated from amino acid content (Perkins, S.J. 1986 Eur. J. Biochem. 157, 169-180). Protein concentrations are also measured by the method of Bradford, M.M. (1976) Anal. Biochem. 72, 248-254, and Lowry, O.H., Rosebrough, N., Farr, A.L. & Randall, R.J. (1951) J. Biol. Chem. 193, pages 265-275, using bovine serum albumin as a standard.

15 SDS-polyacrylamide gels of various concentrations may be purchased from BioRad (Hercules, CA, USA), and stained with Coomassie blue. Molecular weight markers may include rabbit skeletal muscle myosin (200 kDa), *E. coli* (-galactosidase (116 kDa), rabbit muscle phosphorylase B (97.4 kDa), bovine serum albumin (66.2 kDa), ovalbumin (45 kDa), bovine carbonic anhydrase (31 kDa), soybean trypsin inhibitor
 20 (21.5 kDa), egg white lysozyme (14.4 kDa) and bovine aprotinin (6.5 kDa).

EQUIVALENTS

Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments and methods
 25 described herein. The specific embodiments described herein are offered by way of example only, and the invention is to limited only by the terms of the appended claims, along with the full scope of equivalents to which such claims are entitled.

TABLE 2

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT</u> <u>Length</u>	<u>AA</u> <u>Length</u>	<u>Score</u>	<u>Probability</u>
32245287_f2_2	1	5223	139	420	196	1.5e-15
<u>Protein name</u>	<u>Locus Name</u>				<u>Acc#</u>	
hypothetical protein jhp1211	pir:C71832				C71832	
<u>Description</u>						

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT</u> <u>Length</u>	<u>AA</u> <u>Length</u>	<u>Score</u>	<u>Probability</u>
10020167_c1_80	2	5224	611	1836	706	1.4e-69
<u>Protein name</u>	<u>Locus Name</u>				<u>Acc#</u>	
glutaminase A	gp:AB029552				AB029552	
<u>Description</u>						

Aspergillus oryzae gtaA gene for glutaminase A, complete cds.

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT</u> <u>Length</u>	<u>AA</u> <u>Length</u>	<u>Score</u>	<u>Probability</u>
1064765_c1_89	3	5225	249	750	324	1.8e-28
<u>Protein name</u>	<u>Locus Name</u>				<u>Acc#</u>	
alpha-1,6-mannanase	gp:AB024331				AB024331	
<u>Description</u>						

Bacillus circulans aman6 gene for alpha-1,6-mannanase, completecds.

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT</u> <u>Length</u>	<u>AA</u> <u>Length</u>	<u>Score</u>	<u>Probability</u>
10945326_f3_56.....	4	5226	481	1446		
<u>Protein name</u>	<u>Locus Name</u>				<u>Acc#</u>	
<u>Description</u>						

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
12109430_c2_116	5	5227	788	2367	2343	4.6e-243

Protein name

Locus Name

Acc#

immunoreactive 89kD antigen PG87

gp:AF175722

AF175722

Description

Porphyromonas gingivalis strain W50 immunoreactive 89kD antigen PG87 gene, complete cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
14647327_c1_92	6	5228	837	2514	706	2.7e-76

Protein name

Locus Name

Acc#

glutaminase A

gp:AB029552

AB029552

Description

Aspergillus oryzae gtaA gene for glutaminase A, complete cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
19125_c1_98	7	5229	406	1221	760	2.6e-75

Protein name

Locus Name

Acc#

putative aldose 1-epimerase

gp:SC4A7

AL133423

Description

Streptomyces coelicolor cosmid 4A7.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
19536316_c2_107	8	5230	1085	3258	755	1.9e-78

Protein name

Locus Name

Acc#

115K outer membrane protein precursor:SusC protein

pir:JC6027

JC6027

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
22455343_c1_88	9	5231	65	198	55	0.031

Protein name

Locus Name

Acc#

gp:AP000969

AP000969

Description

Oryza sativa genomic DNA, chromosome 1, clone:P0011D01.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
22457686_c2_112	10	5232	724	2175		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
23492786_c3_124	11	5233	85	258		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
23866437_c3_134	12	5234	493	1482	822	6.9e-82

Protein name

Locus Name

Acc#

hypothetical protein SCJ4.42c

pir:T37125

T37125

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24645308_f1_20	13	5235	1207	3624	741	5.9e-71

Protein name

Locus Name

Acc#

hypothetical protein

pir:S76045

S76045

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24647811_c2_108	14	5236	619	1860	126	6.0e-11

Protein name

Locus Name

Acc#

unknown

gp:U96771

U96771

Description

Prevotella bryantii putative polygalacturonase, B-1,4-endoglucanase, and mannanase genes, complete cds; and unknown genes.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
26307018_c2_119	15	5237	427	1284	351	5.6e-32

Protein name

Locus Name

Acc#

sp:YXAH_BACSU

P42107

Description

HYPOTHETICAL 46.2 KD PROTEIN IN ASNH-GNTR INTERGENIC REGION

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
29461537_c3_138.....	16	5238	158	477		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
3027177_c1_87.....	17	5239	336	1011	370	5.4e-34

Protein name

Locus Name

Acc#

endo-arabinase

gp:D85132

D85132

Description

Bacillus subtilis DNA for endo-arabinase, complete cds.

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
4142127_c2_109	18	5240	147	444		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
4492168_c3_125	19	5241	161	486	222	2.3e-17

Protein name

Locus Name

Acc#

alpha-1,6-mannanase

gp:AB024331

AB024331

Description

Bacillus circulans aman6 gene for alpha-1,6-mannanase, completecds.

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
4726687_c1_86	20	5242	240	723		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
818876_c2_110	21	5243	142	429		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
12203438_c1_6	22	5244	420	1263	1627	3.4e-167
<u>Protein name</u>			<u>Locus Name</u>		<u>Acc#</u>	
hypothetical protein			pir:JQ1020		JQ1020	
<u>Description</u>						

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
26365816_f3_4	23	5245	857	2574		
<u>Protein name</u>			<u>Locus Name</u>		<u>Acc#</u>	
<u>Description</u>						

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
32631533_f1_1	24	5246	295	885	435	2.3e-48
<u>Protein name</u>			<u>Locus Name</u>		<u>Acc#</u>	
glucan 1,4-beta-glucosidase, :exo-1,4-beta-glucosidase			pir:JC4825		JC4825	
<u>Description</u>						

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
22275187_f1_1	25	5247	170	513	482	7.4e-46
<u>Protein name</u>			<u>Locus Name</u>		<u>Acc#</u>	
glutathione peroxidase			gp:LLAJ109		AJ000109	
<u>Description</u>						

Lactococcus lactis carB and gpo genes.

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
24432962_f2_1	26	5248	91	273		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
3.176.706.7_f1_1	27	5249	294	885		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
14257180_f2_19	28	5250	493	1482	250	6.5e-36

Protein name

Locus Name

Acc#

sp:ARSF_HUMAN

P54793

Description

ARYLSULFATASE F PRECURSOR, (ASF)

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
24025282_c2_75	29	5251	550	1653	1045	1.6e-105

Protein name

Locus Name

Acc#

sp:HEXA_FORGI

P49008

Description

(BETA-NAHASE)

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24486016_f2_21	30	5252	980	2943	165	1.4e-09
Protein name			Locus Name		Acc#	
response regulator			gp:SPAJ6398		AJ006398	
Description						
Streptococcus pneumoniae rr09 and hk09 genes; two component system09.						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
25584525_f1_10	31	5253	786	2361	212	2.8e-14
Protein name			Locus Name		Acc#	
putative secreted protein			gp:SCF41		AL117387	
Description						
Streptomyces coelicolor cosmid F41.						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
26209530_c3_98	32	5254	535	1608	280	5.0e-32
Protein name			Locus Name		Acc#	
phosphonate monoester hydrolase			gp:BCU44852		U44852	
Description						
Burkholderia caryophylli PG2982 phosphonate monoester hydrolase(pehA) gene, complete cds.						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
267517_f3_30	33	5255	60	183		
Protein name			Locus Name		Acc#	
Description						
NO-HIT						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
2932812_c2_68	34	5256	509	1530	258	1.4e-34

Protein name

Locus Name

Acc#

sp:ARSE_HUMAN

P51690

Description

ARYLSULFATASE E PRECURSOR, (ASE)

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
3020203_c2_67	35	5257	423	1272	666	2.3e-65

Protein name

Locus Name

Acc#

sp:HEXA_PORGI

P49008

Description

(BETA-NAHASE)

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
3949012_c2_69	36	5258	487	1464	726	1.0e-71

Protein name

Locus Name

Acc#

sp:MODF_ECOLI

P31060

Description

PROTEIN PHRA)

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4726588_c3_86	37	5259	370	1113	1233	1.9e-125

Protein name

Locus Name

Acc#

hypothetical protein b2097

pir:H64976

H64976

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
5273452_c1_61	38	5260	248	747	783	9.4e-78

Protein name

Locus Name

Acc#

sp:PMGI_ECOLI

P31217

Description

(PGAM 1) (BPG-DEPENDENT PGAM 1)

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
10569587_c1_59	39	5261	768	2307	924	1.1e-92

Protein name

Locus Name

Acc#

melibiase

gp:TEMELA

Y08557

Description

T.ethanolicus melA and lacA genes.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
1369077_f3_49	40	5262	405	1218		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
15631576_c1_74	41	5263	400	1203	111	0.0016

Protein name

Locus Name

Acc#

cytochrome-c oxidase, chain III

pir:S36954

S36954

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
16532750_c2_104	42	5264	174	525	140	3.5e-08

Protein name

Locus Name

Acc#

F14N23.29

gp:AC005489

AC005489

Description

Genomic sequence for Arabidopsis thaliana BAC F14N23 from Chromosome 1, complete sequence.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
16828427_f1_5	43	5265	422	1269	593	1.3e-57

Protein name

Locus Name

Acc#

N utilization substance protein A

pir:H72213

H72213

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
19535126_f3_53	44	5266	284	855	820	1.1e-81

Protein name

Locus Name

Acc#

sp:ABCX_CYAPA

P48255

Description

PROBABLE ATP-DEPENDENT TRANSPORTER YCF16

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
20214450_f1_7	45	5267	64	195		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
2054635_f2_34	46	5268	115	348	353	3.4e-32
Protein name			Locus Name		Acc#	
hypothetical protein b0866			pir:B64825		B64825	
Description						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
20724002_c1_63	47	5269	553	1662	742	2.1e-73
Protein name			Locus Name		Acc#	
probable secreted alpha-galactosidase			pir:T36472		T36472	
Description						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
22519677_f1_3	48	5270	449	1350	1532	4.0e-157
Protein name			Locus Name		Acc#	
L-fucose permease			gp:AF137263		AF137263	
Description						
Bacteroides thetaiotaomicron 30S ribosomal protein S16-like protein, fucose gene cluster, and RNA polymerase sigma factor SigZ-like protein (sigZ) genes, complete cds.						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
22917183_c1_62	49	5271	573	1722	297	9.3e-23
Protein name			Locus Name		Acc#	
receptor antigen (RagA)			gp:PGI130872		AJ130872	
Description						
Porphyromonas gingivalis W50 receptor antigen (rag) locus encoding a major immunodominant 55kDa antigen.						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
23679512_c1_61	50	5272	555	1668	432	7.7e-42

Protein name

Locus Name

Acc#

115K outer membrane protein precursor:SusC protein

pir:JC6027

JC6027

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24490677_c3_119	51	5273	136	411	111	1.6e-06

Protein name

Locus Name

Acc#

probable sigK protein

pir:F70830

F70830

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24640927_c3_121	52	5274	616	1851	116	0.00016

Protein name

Locus Name

Acc#

unknown

gp:U96771

U96771

Description

Prevotella bryantii putative polygalacturonase, B-1,4-endoglucanase, and mannanase genes, complete cds; and unknown genes.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24803457_c1_65	53	5275	112	339		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
25963962_f2_28	54	5276	377	1134	123	0.00063

Protein name

Locus Name

Acc#

sp:TRHY_RABIT

P37709

Description

TRICHOHYALIN

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
26212805_f2_29	55	5277	156	471	119	2.2e-07

Protein name

Locus Name

Acc#

sp:YHBC_ECOLI

P03843

Description

HYPOTHETICAL 16.8 KD PROTEIN IN NUSA-METY INTERGENIC REGION

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
26364056_f2_31	56	5278	521	1566	1803	7.7e-186

Protein name

Locus Name

Acc#

sp:Y074_SYNY3

Q55790

Description

HYPOTHETICAL 52.8 KD PROTEIN SLR0074

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
29480306_f3_51	57	5279	81	246		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
30578126_f1_4	58	5280	446	1341		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
31663925_f3_50.....	59	5281	89	270	76	0.0077

Protein name

Locus Name

Acc#

probable serine proteinase

pir:T36552

T36552

Description

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
33242938_f2_33.....	60	5282	456	1371	433	1.1e-40

Protein name

Locus Name

Acc#

sp:Y076_SYNY3

Q55792

Description

HYPOTHETICAL 50.0 KD PROTEIN SLR0076

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
33448342_f1_8.....	61	5283	253	762		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
35442313_f1_1	62	5284	292	879	864	2.4e-86

Protein name

Locus Name

Acc#

sp:FUCO_ECOLI P11549

Description

LACTALDEHYDE REDUCTASE, (PROPANEDIOL OXIDOREDUCTASE)

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
3942813_f3_47	63	5285	215	648	1026	1.7e-103

Protein name

Locus Name

Acc#

L-fucose-1-phosphate aldolase gp:AF137263 AF137263

Description

Bacteroides thetaiotaomicron 30S ribosomal protein S16-likeprotein, fucose gene cluster, and RNA polymerase sigma factorSigZ-like protein (sigZ) genes, complete cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4001515_c1_60.....	64	5286	388	1167	135	3.7e-06

Protein name

Locus Name

Acc#

transmembrane sensor gp:AF051691 AF051691

Description

Pseudomonas aeruginosa stress factor A (psfA), ECF sigma factor(fiuI), transmembrane sensor (fiuR), and hydroxamate-typeferrisiderophore receptor (fiuA) genes, complete cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4196001_f2_26.....	65	5287	488	1467	1748	5.2e-180

Protein name

Locus Name

Acc#

L-fucose kinase gp:AF137263 AF137263

Description

Bacteroides thetaiotaomicron 30S ribosomal protein S16-likeprotein, fucose gene cluster, and RNA polymerase sigma factorSigZ-like protein (sigZ) genes, complete cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
5197137_f1_6	66	5288	1016	3051	1669	1.2e-171

Protein name

Locus Name

Acc#

Initiation factor IF2-alpha

gp:ECAJ2540

AJ002540

Description

Escherichia coli (strain EcoAU9307) infB gene encoding translational initiation factor IF2.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
5366453_f1_9	67	5289	416	1251	1106	5.5e-112

Protein name

Locus Name

Acc#

nifS-like protein

gp:MLCB22

Z98741

Description

Mycobacterium leprae cosmid B22.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
6252033_f2_27	68	5290	67	204	259	8.1e-22

Protein name

Locus Name

Acc#

L-fucose permease

gp:AF137263

AF137263

Description

Bacteroides thetaiotaomicron 30S ribosomal protein S16-like protein, fucose gene cluster, and RNA polymerase sigma factor SigZ-like protein (sigZ) genes, complete cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
10742010_c1_61	69	5291	695	2088	633	7.3e-62

Protein name

Locus Name

Acc#

sp:PFLD_ECOLI

P32674

Description

FORMATE ACETYLTRANSFERASE 2, (PYRUVATE FORMATE-LYASE 2)

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
13913887_c2_82	70	5292	694	2085	564	2.7e-74
Protein name			Locus Name		Acc#	
hypothetical protein TM0280			pir:F72395		F72395	
Description						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
15660937_c3_108	71	5293	260	783		
Protein name			Locus Name		Acc#	
Description						

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
20734637_c2_80	72	5294	263	792	306	3.9e-45
Protein name			Locus Name		Acc#	
probable pyruvate formate-lyase activating enzyme, pflC homolog			pir:A69431		A69431	
Description						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
21516933_f1_15	73	5295	76	231		
Protein name			Locus Name		Acc#	
Description						

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
22461007_f3_37	74	5296	235	708	241	2.5e-20
Protein name			Locus Name		Acc#	
probable competence protein ComF			pir:F75402		F75402	
Description						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
23437627_f1_1	75	5297	88	267		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
29350677_c2_79	76	5298	1368	4107	468	1.7e-40

Protein name

Locus Name

Acc#

bZIP histidine kinase

gp:PPUY18245

Y18245

Description

Pseudomonas putida todX, todF, todC1, todC2, todB, todA, todD, todE, todG, todI, todH, todS, todT genes.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
30573761_f1_4	77	5299	69	210		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
36225337_f2_19	78	5300	382	1149		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
36585962_f3_35	79	5301	418	1257	722	2.7e-71

Protein name	Locus Name	Acc#
alpha galactosidase precursor	gp:AF061331	AF061331

Description

Saccharopolyspora erythraea alpha galactosidase precursor (melA) gene, complete cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
3940877_c1_67	80	5302	96	291	163	4.7e-12

Protein name	Locus Name	Acc#
	sp:YCNE_BACSU	P94425

Description

HYPOTHETICAL 10.9 KD PROTEIN IN PHRC-GDH INTERGENIC REGION

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4094627_f1_17.....	81	5303	1060	3183	879	6.3e-88

Protein name	Locus Name	Acc#
115K outer membrane protein precursor:SusC protein	pir:JC6027	JC6027

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4101507_c3_101.....	82	5304	382	1149	842	5.2e-84

Protein name	Locus Name	Acc#
putative aldose 1-epimerase	gp:SC4A7	AL133423

Description

Streptomyces coelicolor cosmid 4A7.

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
4103812_f2_21	83	5305	271	816	360	6.2e-33

Protein name

Locus Name

Acc#

sp:SUHB_ECOLI

Description

EXTRAGENIC SUPPRESSOR PROTEIN SUHB

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
4422768_f1_18	84	5306	95	288		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
4534660_c1_62	85	5307	443	1332	620	5.5e-78

Protein name

Locus Name

Acc#

sp:XYLE_ECOLI

P09098

Description

D-XYLOSE-PROTON SYMPORTER (D-XYLOSE TRANSPORTER)

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
4799033_f2_30	86	5308	65	198		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
10820130_c1_218	87	5309	68	207		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
11723437_c3_354	88	5310	79	240		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
1207152_c3_351	89	5311	162	489	217	8.9e-18

Protein name

Locus Name

Acc#

unknown

gp:AF125164

AF125164

Description

Bacteroides fragilis 638R polysaccharide B (PS B2) biosynthesis locus, complete sequence; and unknown genes.

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
12754816_f2_114	90	5312	209	630		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
14296885_c1_248	91	5313	200	603	260	2.5e-22
<u>Protein name</u>			<u>Locus Name</u>			<u>Acc#</u>
conserved hypothetical protein AF0781			pir:E69347			E69347
<u>Description</u>						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
14626317_f1_52	92	5314	640	1923	736	1.0e-78
<u>Protein name</u>			<u>Locus Name</u>			<u>Acc#</u>
			gp:ECU89166			U89166
<u>Description</u>						
Eikenella corrodens lysine decarboxylase (ECORLD) gene, complete cds.						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
14648562_f2_116	93	5315	579	1740	889	5.5e-89
<u>Protein name</u>			<u>Locus Name</u>			<u>Acc#</u>
single-strand DNA-specific exonuclease homolog yrvE			pir:H69980			H69980
<u>Description</u>						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
15735882_c1_208	94	5316	400	1203	696	1.5e-68
<u>Protein name</u>			<u>Locus Name</u>			<u>Acc#</u>
renin-binding protein-related protein:protein slr1975:protein slr1975			pir:S75649			S75649
<u>Description</u>						

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
16016075_c2_288	95	5317	85	258		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
16616302_f3_197.....	96	5318	439	1320	1138	2.3e-115

Protein name

Locus Name

Acc#

coenzyme F390 synthetase (ftsA-3) homolog

pir:D69501

D69501

Description

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
16835312_f1_34.....	97	5319	279	840		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
192693_c2_289.....	98	5320	211	636	236	1.4e-19

Protein name

Locus Name

Acc#

sp:YHCG_ECOLI

P45423

Description

HYPOTHETICAL 43.3 KD PROTEIN IN GLTF-NANT INTERGENIC REGION (O375)

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	--------------	--------------	-------	-------------

19535652_c3_316	99	5321	640	1923	940	2.2e-94
-----------------	----	------	-----	------	-----	---------

Protein name

Locus Name

Acc#

putative epimerase/dehydratase WbiI

gp:AF064070

AF064070

Description

Burkholderia pseudomallei putative dihydroorotase (pyrC) gene, partial cds; putative 1-acyl-sn-glycerol-3-phosphateacyltransferase (plsC), putative diadenosine tetrphosphatase (apaH), complete cds; type II O-antigen biosynthesis gene cluster, complete sequence; putative undecaprenyl phosphateN-acetylglucosaminyltransferase, and putative

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	--------------	--------------	-------	-------------

2031552_c2_298	100	5322	101	306	74	0.013
----------------	-----	------	-----	-----	----	-------

Protein name

Locus Name

Acc#

sp:NU3M_RAT

P05506

Description

NADH-UBIQUINONE OXIDOREDUCTASE CHAIN 3,

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	--------------	--------------	-------	-------------

20516500_c2_304	101	5323	705	2118		
-----------------	-----	------	-----	------	--	--

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	--------------	--------------	-------	-------------

211687_c2_299	102	5324	461	1386	309	2.0e-27
---------------	-----	------	-----	------	-----	---------

Protein name

Locus Name

Acc#

Cap8K

gp:SAU73374

U73374

Description

Staphylococcus aureus type 8 capsule genes, cap8A, cap8B, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds.

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
2117305_c3_317	103	5325	856	2571	464	2.7e-56

Protein name

Locus Name

Acc#

otnA protein

pir:S70958

S70958

Description

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
21485952_c1_244.....	104	5326	965	2898	821	1.0e-120

Protein name

Locus Name

Acc#

sp:YDIJ_ECOLI

P77748

Description

HYPOTHETICAL 113.2 KD PROTEIN IN LPP-AROD INTERGENIC REGION

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
21500933_f1_41.....	105	5327	1084	3255		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
22345000_f2_88.....	106	5328	88	267		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
22551562_f1_48	107	5329	143	432		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
23546952_f2_106	108	5330	291	876		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
23567187_f2_102	109	5331	109	330		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
23609457_f1_27	110	5332	64	195		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
23633592_c1_237	111	5333	64	195		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
-----------------	-------------	-------------	------------------	------------------	--------------	--------------------

24103388_f1_71	112	5334	531	1596	831	7.7e-83
----------------	-----	------	-----	------	-----	---------

Protein name

Locus Name

Acc#

indolepyruvate oxidoreductase, alpha subunit

pir:G69114

G69114

Description

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
-----------------	-------------	-------------	------------------	------------------	--------------	--------------------

24303127_f1_33	113	5335	142	429		
----------------	-----	------	-----	-----	--	--

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
-----------------	-------------	-------------	------------------	------------------	--------------	--------------------

24318802_f3_187	114	5336	410	1233	490	1.0e-46
-----------------	-----	------	-----	------	-----	---------

Protein name

Locus Name

Acc#

sp:XYLR_ANATH

Q44406

Description

XYLOSE REPRESSOR

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
-----------------	-------------	-------------	------------------	------------------	--------------	--------------------

24402177_f2_103	115	5337	94	285		
-----------------	-----	------	----	-----	--	--

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

24415908_f1_72	116	5338	195	588	310	1.2e-27
----------------	-----	------	-----	-----	-----	---------

Protein name

Locus Name

Acc#

indolepyruvate ferredoxin oxidoreductase,
subunit beta (iorB) homolog

pir:E69503

E69503

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

24415930_c1_239	117	5339	322	969	937	4.5e-94
-----------------	-----	------	-----	-----	-----	---------

Protein name

Locus Name

Acc#

WbpB

gp:PAU50396

U50396

Description

Pseudomonas aeruginosa Wzz (Rol) (wzz (rol)) gene, partial cds, WbpA (wbpB), WbpB (wbpB), WbpC (wbpC), WbpD (wbpD), WbpE (wbpE), Wzy (Rfc) (wzy (rfc)), Wzx (wzx), HisH (hisH), HisF (hisF), WbpG (wbpG), WbpH (wbpH), WbpI (wbpI), WbpJ (wbpJ), WbpK (wbpK), WbpL (wbpL), WbpM (wbpM) and WbpN (wbpN) genes, complete cds, and UvrB (uvrB) gene, partial cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

24426337_c1_216	118	5340	89	270		
-----------------	-----	------	----	-----	--	--

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

24640875_c1_245	119	5341	376	1131	1914	1.3e-197
-----------------	-----	------	-----	------	------	----------

Protein name

Locus Name

Acc#

putative aminotransferase

gp:AF125164

AF125164

Description

Bacteroides fragilis 638R polysaccharide B (PS B2) biosynthesis locus, complete sequence; and unknown genes.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24651562_f1_68	120	5342	522	1569	318	3.1e-25
Protein name			Locus Name			Acc#
surface antigen BspA			pir:T31094			T31094
Description						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24806502_c1_249.....	121	5343	325	978	581	2.4e-56
Protein name			Locus Name			Acc#
			sp:FMT_BACSU			
Description						
METHIONYL-TRNA FORMYLTRANSFERASE,						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
25421942_c1_207.....	122	5344	181	546	152	6.9e-11
Protein name			Locus Name			Acc#
unknown			gp:AF048749			AF048749
Description						
Bacteroides fragilis capsular polysaccharide biosynthesis operon, complete sequence.						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
25429812_f3_171.....	123	5345	341	1026	111	0.0022
Protein name			Locus Name			Acc#
			sp:Y973_METJA			Q58383
Description						
HYPOTHETICAL PROTEIN MJ0973						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

25583577_c3_356	124	5346	393	1182	172	3.4e-10
-----------------	-----	------	-----	------	-----	---------

Protein name

Locus Name

Acc#

Cap5I

gp:SAU81973

U81973

Description

Staphylococcus aureus capsule gene cluster Cap5A through Cap5Pgenes, complete cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

25667592_c3_363	125	5347	606	1821	537	4.0e-55
-----------------	-----	------	-----	------	-----	---------

Protein name

Locus Name

Acc#

chloride channel, probable, homolog

pir:F69426

F69426

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

25676387_c3_352	126	5348	83	252	118	2.8e-07
-----------------	-----	------	----	-----	-----	---------

Protein name

Locus Name

Acc#

tachylectin-3

gp:AB017484

AB017484

Description

Tachypleus tridentatus mRNA for tachylectin-3, complete cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

25976510_c2_263	127	5349	311	936	271	1.7e-23
-----------------	-----	------	-----	-----	-----	---------

Protein name

Locus Name

Acc#

gp:ECNPL

X03345

Description

E. coli npl gene for N-acetylneuraminate lyase subunit (EC4.1.3.3).

25583577_c3_356

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

26604635_c3_355	128	5350	370	1113	678	1.3e-66
-----------------	-----	------	-----	------	-----	---------

Protein name

Locus Name

Acc#

unknown

gp:AF144879

AF144879

Description

Leptospira interrogans rfb locus, complete sequence.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

29305313_c1_242	129	5351	490	1473	122	0.00028
-----------------	-----	------	-----	------	-----	---------

Protein name

Locus Name

Acc#

putative polysaccharide polymerase

gp:SPU09239

U09239

Description

Streptococcus pneumoniae type 19F capsular polysaccharide biosynthesis operon, (cps19fABCDEFGHIJKLMNO) genes, complete cds, and aliA gene, partial cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

29317660_c1_204	130	5352	125	378		
-----------------	-----	------	-----	-----	--	--

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

29509630_c1_217	131	5353	70	213		
-----------------	-----	------	----	-----	--	--

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
30355312_c1_246	132	5354	165	498	102	1.4e-05
Protein name			Locus Name			Acc#
DNA-binding protein HB			pir:C75600			C75600
Description						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
31381_c1_209	133	5355	416	1251	289	6.5e-24
Protein name			Locus Name			Acc#
			sp:YYBO_BACSU			P37489
Description						
HYPOTHETICAL 48.2 KD PROTEIN IN COTF-TETB INTERGENIC REGION						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
31414025_c2_290	134	5356	227	684	505	2.7e-48
Protein name			Locus Name			Acc#
			sp:3MG1_ECOLI			P05100
Description						
I)						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
31882036_c2_271	135	5357	73	222		
Protein name			Locus Name			Acc#
Description						
NO-HIT						

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
32064137_c2_294	136	5358	209	630	120	2.3e-07

Protein name

Locus Name

Acc#

unknown

gp:AF048749

AF048749

Description

Bacteroides fragilis capsular polysaccharide biosynthesis operon, complete sequence.

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
32323912_f1_17	137	5359	411	1236		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
32429512_f2_113.....	138	5360	771	2316	163	1.1e-18

Protein name

Locus Name

Acc#

DNA repair protein RAD25 homolog

pir:F69294

F69294

Description

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
32617177_f2_95.....	139	5361	92	279		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
-----------------	-------------	-------------	------------------	------------------	--------------	--------------------

337_c3_353	140	5362	198	597	532	3.7e-51
------------	-----	------	-----	-----	-----	---------

Protein name

Locus Name

Acc#

acetyl transferase homolog

pir:S70673

S70673

Description

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
-----------------	-------------	-------------	------------------	------------------	--------------	--------------------

34011402..f1..20.....	141	5363	74	225		
-----------------------	-----	------	----	-----	--	--

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
-----------------	-------------	-------------	------------------	------------------	--------------	--------------------

34066312..c1..238.....	142	5364	177	534		
------------------------	-----	------	-----	-----	--	--

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
-----------------	-------------	-------------	------------------	------------------	--------------	--------------------

34116277..f3..191.....	143	5365	60	183	115	5.7e-07
------------------------	-----	------	----	-----	-----	---------

Protein name

Locus Name

Acc#

hypothetical protein 3

pir:S28487

S28487

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
34167567_c2_297	144	5366	446	1341	1226	1.1e-124

Protein name

Locus Name

Acc#

ORF1P

gp:AB025970

AB025970

Description

Plesiomonas shigelloides gene for ORF1P, ORF2P, ORF3P, ORF4P, ORF5P, ORF6P, ORF7P, ORF8P, ORF9P, ORF10P, ORF11P.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
36361063_c2_300	145	5367	370	1113	610	2.0e-59

Protein name

Locus Name

Acc#

WbpH

gp:PAU50396

U50396

Description

Pseudomonas aeruginosa Wzz (Rol) (wzz (rol)) gene, partial cds, WbpA (wbpB), WbpB (wbpB), WbpC (wbpC), WbpD (wbpD), WbpE (wbpE), Wzy (Rfc) (wzy (rfc)), Wzx (wzx), HisH (hisH), HisF (hisF), WbpG (wbpG), WbpH (wbpH), WbpI (wbpI), WbpJ (wbpJ), WbpK (wbpK), WbpL (wbpL), WbpM (wbpM) and WbpN (wbpN) genes, complete cds, and UvrB (uvrB) gene, partial cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
3911015_c3_358.....	146	5368	180	543	153	5.4e-11

Protein name

Locus Name

Acc#

serine O-acetyltransferase,

pir:E53402

E53402

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
3948587_c2_296.....	147	5369	399	1200	1198	9.9e-122

Protein name

Locus Name

Acc#

gp:D64132

D64132

Description

Porphyromonas gingivalis PorR and PorS genes, complete cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4022178_c2_262	148	5370	373	1122	86	7.1e-07

Protein name

Locus Name

Acc#

sp:YCCC_ECOLI

Description

HYPOTHETICAL 81.2 KD PROTEIN IN APPA-CSPH INTERGENIC REGION

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4033125_c1_250	149	5371	226	681	546	1.2e-52

Protein name

Locus Name

Acc#

ribulose-5-phosphate 3-epimerase homolog ylor

pir:B69879

B69879

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4065677_f2_133	150	5372	348	1047	283	9.0e-25

Protein name

Locus Name

Acc#

conserved hypothetical protein BB0709

pir:D70188

D70188

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4096877_f1_69	151	5373	296	891	348	1.2e-31

Protein name

Locus Name

Acc#

sp:NUC_BORBU

051372

Description

PUTATIVE ENDONUCLEASE BB0411,

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

4103375_c1_243	152	5374	410	1233	843	4.1e-84
----------------	-----	------	-----	------	-----	---------

Protein name

Locus Name

Acc#

putative transferase

gp:BBR007747

AJ007747

Description

Bordetella bronchiseptica cosmid BbLPS1.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

4457512_f3_184	153	5375	333	1002	221	2.6e-16
----------------	-----	------	-----	------	-----	---------

Protein name

Locus Name

Acc#

conserved hypothetical protein MTH83

pir:F69210

F69210

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

4459660_c2_252	154	5376	269	810	109	9.0e-06
----------------	-----	------	-----	-----	-----	---------

Protein name

Locus Name

Acc#

probable NADH-plastoquinone oxidoreductase subunit

pir:C71018

C71018

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

4728385_f3_173	155	5377	724	2175	208	6.3e-13
----------------	-----	------	-----	------	-----	---------

Protein name

Locus Name

Acc#

probable purine NTPase PAB0812

pir:F75103

F75103

Description

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
-----------------	-------------	-------------	------------------	------------------	--------------	--------------------

4786250_f2_83	156	5378	156	471	135	4.3e-09
---------------	-----	------	-----	-----	-----	---------

Protein name

Locus Name

Acc#

hypothetical protein MTH658

pir:E69187

E69187

Description

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
-----------------	-------------	-------------	------------------	------------------	--------------	--------------------

4789066_f2_104	157	5379	71	216		
----------------	-----	------	----	-----	--	--

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
-----------------	-------------	-------------	------------------	------------------	--------------	--------------------

5359842_c3_319	158	5380	174	525		
----------------	-----	------	-----	-----	--	--

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
-----------------	-------------	-------------	------------------	------------------	--------------	--------------------

5877042_c1_251	159	5381	325	975	131	3.0e-05
----------------	-----	------	-----	-----	-----	---------

Protein name

Locus Name

Acc#

sp:CME3_BACSU

P39695

Description

COME OPERON PROTEIN 3

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

5894001_f1_46	160	5382	83	252	65	0.020
---------------	-----	------	----	-----	----	-------

Protein name

Locus Name

Acc#

sp:UDG_STRPY

Q07172

Description

(UDP-GLCDH) (UDPGDH)

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

6056562_c3_359	161	5383	204	615	406	8.3e-38
----------------	-----	------	-----	-----	-----	---------

Protein name

Locus Name

Acc#

putative transferase

gp:BBR007747

AJ007747

Description

Bordetella bronchiseptica cosmid BbLPS1.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

6288313_f2_105.....	162	5384	308	927	407	6.5e-38
---------------------	-----	------	-----	-----	-----	---------

Protein name

Locus Name

Acc#

transposase

gp:AF038866

AF038866

Description

Bacteroides fragilis transposon Tn5520 transposase (bipH) and mobilization protein BmpH (bmpH) genes, complete cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

781512_f2_123.....	163	5385	82	249		
--------------------	-----	------	----	-----	--	--

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
962777_f1_73	164	5386	190	573	457	3.3e-43

Protein name

Locus Name

Acc#

sp:XPT_BACSU

P42085

Description

XANTHINE PHOSPHORIBOSYLTRANSFERASE,

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
9765913_c1_232	165	5387	67	204	75	0.013

Protein name

Locus Name

Acc#

sp:HBB_PANPO

P04244

Description

HEMOGLOBIN BETA CHAIN

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
10671885_c2_125.....	166	5388	123	372		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
10735927_c2_130.....	167	5389	335	1008	124	3.2e-07

Protein name

Locus Name

Acc#

actinorhodin polyketide dimerase-related protein

pir:C72410

C72410

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
10757837_c1_119	168	5390	415	1248	404	1.4e-37

Protein name

Locus Name

Acc#

sp:YRKO_BACSU

P54442

Description

HYPOTHETICAL 46.4 KD PROTEIN IN BLTR-SPOIIC INTERGENIC REGION

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
11953533_f1_20	169	5391	75	228	69	0.0020

Protein name

Locus Name

Acc#

sp:HXD3_BRARE

O42370

Description

HOMEBOX PROTEIN HOX-D3 (FRAGMENT)

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
12541502_f2_51	170	5392	78	237		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
12693752_f2_64	171	5393	208	627	451	1.4e-42

Protein name

Locus Name

Acc#

putative GTP-binding protein

gp:ATAC004786

AC004786

Description

Arabidopsis thaliana chromosome II BAC T20K9 genomic sequence, complete sequence.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

14642187_f2_61	172	5394	175	528	511	6.2e-49
----------------	-----	------	-----	-----	-----	---------

Protein name

Locus Name

Acc#

sp:Y318_HAEIN

P43984

Description

HYPOTHETICAL PROTEIN HI0318

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

15125662_c3_164	173	5395	495	1488	367	1.6e-33
-----------------	-----	------	-----	------	-----	---------

Protein name

Locus Name

Acc#

gp:D90837

Description

E.coli genomic DNA, Kohara clone #347(44.2-44.5 min.).

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

16907752_c3_168	174	5396	360	1083	973	6.9e-98
-----------------	-----	------	-----	------	-----	---------

Protein name

Locus Name

Acc#

sp:YODE_PSEAE

Q01609

Description

HYPOTHETICAL 40.7 KD PROTEIN IN OPDE 3'REGION (ORF2)

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

20579675_c1_104	175	5397	213	642	496	2.4e-47
-----------------	-----	------	-----	-----	-----	---------

Protein name

Locus Name

Acc#

recR protein

pir:H75547

H75547

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
20587753_c3_165	176	5398	388	1167	882	3.0e-88

Protein name

Locus Name

Acc#

sp:PATB_BACSU

Q08432

Description

PUTATIVE AMINOTRANSFERASE B,

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
20601437_f2_65	177	5399	204	615		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
21648312_f2_62	178	5400	109	330		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
22853452_c3_156	179	5401	179	540	158	1.6e-11

Protein name

Locus Name

Acc#

sp:YP20_BACLI

P05332

Description

HYPOTHETICAL P20 PROTEIN

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
22867327_c3_155	180	5402	149	450		

Protein name _____ Locus Name _____ Acc# _____

Description _____

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
23492786_f3_67	181	5403	70	213		

Protein name _____ Locus Name _____ Acc# _____

Description _____

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
23945302_c2_140	182	5404	230	693	163	4.7e-12

Protein name _____ Locus Name _____ Acc# _____

sp:RIBD_METJA Q58085

Description _____

PUTATIVE RIBOFLAVIN BIOSYNTHESIS ENZYME

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24334637_c1_122	183	5405	406	1221	223	3.6e-16

Protein name _____ Locus Name _____ Acc# _____

cation efflux system (czcB-like) pir:E70342 E70342

Description _____

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24416552_c1_118	184	5406	123	372	224	1.6e-18

Protein name _____ Locus Name _____ Acc# _____

oxidoreductase, aldo/keto reductase family pir:H72307 H72307

Description _____

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
26594087_c1_109	185	5407	379	1140	896	9.9e-90

Protein name	Locus Name	Acc#
oxidoreductase, aldo/keto reductase family	pir:H72307	H72307
Description		

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
3237755_c2_132	186	5408	225	678	376	1.3e-34

Protein name	Locus Name	Acc#
plant-metabolite dehydrogenase homolog yvgN	pir:C70040	C70040
Description		

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
33392187_c1_120	187	5409	342	1029	608	3.3e-59

Protein name	Locus Name	Acc#
oxidoreductase, aldo/keto reductase family	pir:H72307	H72307
Description		

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
34198387_c2_134	188	5410	292	879		

Protein name	Locus Name	Acc#
Description		

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
35313816_c1_111	189	5411	287	864	761	2.0e-75

Protein name	Locus Name	Acc#
oxidoreductase, aldo/keto reductase family	pir:A72308	A72308
Description		

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
3948575_c2_123	190	5412	227	684	584	1.1e-56

Protein name

Locus Name

Acc#

sp:YF08_METJA

Q58903

Description

HYPOTHETICAL ABC TRANSPORTER ATP-BINDING PROTEIN MJ1508

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4064178_f1_25	191	5413	455	1368	197	8.2e-13

Protein name

Locus Name

Acc#

aspartate aminotransferase

gp:AF035157

AF035157

Description

Lactococcus lactis aspartate aminotransferase (aspC) gene, completecds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4335425_f3_98	192	5414	498	1497	370	5.4e-34

Protein name

Locus Name

Acc#

hypothetical protein

pir:S75887

S75887

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4410135_f2_56	193	5415	141	426		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
4486261_c1_121	194	5416	164	495		

Protein name Locus Name Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
4589627_f3_93	195	5417	199	600	257	5.1e-22

Protein name Locus Name Acc#

yqge hypothetical protein pir:H72114 H72114

Description

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
47157165_c1_100	196	5418	75	228		

Protein name Locus Name Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
632012_f1_22	197	5419	82	249		

Protein name Locus Name Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
959687_f1_28	198	5420	140	423		

Protein name Locus Name Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
10631882_c2_238	199	5421	61	186		

Protein name Locus Name Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
11148453_F3_122	200	5422	127	384	126	3.9e-08

Protein name Locus Name Acc#

hypothetical protein yngA pir:F69892 F69892

Description

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
12381962_F3_134	201	5423	1058	3177	2370	6.3e-246

Protein name Locus Name Acc#

hypothetical protein mexF pir:T30830 T30830

Description

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
1256327_G3_274	202	5424	149	450	96	1.6e-06

Protein name Locus Name Acc#

ct469 hypothetical protein pir:D72060 D72060

Description

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
12690817_c2_225	203	5425	141	426		

Protein name Locus Name Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
12712827_f3_120	204	5426	353	1062	276	5.0e-24
Protein name				Locus Name		Acc#
conserved hypothetical protein				pir:F72386		F72386
Description						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
12922202_f3_116	205	5427	316	951	132	3.0e-06
Protein name				Locus Name		Acc#
hypothetical protein aq_380				pir:A70334		A70334
Description						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
1370468_f3_118	206	5428	93	282		
Protein name				Locus Name		Acc#
Description				NO-HIT		

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
13878425_f2_72	207	5429	134	405	368	8.9e-34
Protein name				Locus Name		Acc#
				sp:YYAH_BACSU		P37516
Description						
HYPOTHETICAL 14.4 KD PROTEIN IN TETB-EXOA INTERGENIC REGION (ORFF)						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
14508425_c1_192	208	5430	236	711	166	2.3e-12
Protein name				Locus Name		Acc#
hypothetical protein MTH939				pir:G69225		G69225
Description						

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT</u> <u>Length</u>	<u>AA</u> <u>Length</u>	<u>Score</u>	<u>Probability</u>
14646956_c2_261	209	5431	82	249	61	0.023

Protein name

Locus Name

Acc#

mannanase

gp:U96771

U96771

Description

Prevotella bryantii putative polygalacturonase, B-1,4-endoglucanase, and mannanase genes, complete cds; and unknowngenes.

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT</u> <u>Length</u>	<u>AA</u> <u>Length</u>	<u>Score</u>	<u>Probability</u>
14705261_f1_15	210	5432	501	1506	115	0.00082

Protein name

Locus Name

Acc#

unknown protein

gp:BACCOMGA

Description

Bacillus subtilis (clone pED4) comG-(1,2,3,4,5,6,and 7) proteins incomG operon, complete cds.

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT</u> <u>Length</u>	<u>AA</u> <u>Length</u>	<u>Score</u>	<u>Probability</u>
14744010_f3_129.....	211	5433	265	798	275	6.3e-24

Protein name

Locus Name

Acc#

conserved hypothetical protein yjka

pir:E69851

E69851

Description

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT</u> <u>Length</u>	<u>AA</u> <u>Length</u>	<u>Score</u>	<u>Probability</u>
14876313_f2_84.....	212	5434	667	2004	1443	1.1e-147

Protein name

Locus Name

Acc#

DNA ligase

gp:BST011676

AJ011676

Description

Bacillus stearothermophilus lig gene.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
15641902_c3_288	213	5435	202	609	345	2.4e-31
Protein name			Locus Name		Acc#	
conserved hypothetical protein AF2201			pir:A69525		A69525	
Description						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
16204652_f2_65	214	5436	604	1815	134	3.1e-06
Protein name			Locus Name		Acc#	
hypothetical protein AF1867			pir:B69483		B69483	
Description						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
16828382_f3_137	215	5437	305	918	801	1.2e-79
Protein name			Locus Name		Acc#	
			sp:PYRD_AQUAE		066461	
Description						
(DHODEHASE)						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
16832885_c3_307	216	5438	431	1296	1724	1.8e-177
Protein name			Locus Name		Acc#	
hypothetical protein			pir:JQ1020		JQ1020	
Description						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
188752_f1_20	217	5439	761	2286	380	3.5e-34
Protein name			Locus Name		Acc#	
conserved hypothetical protein AF1878			pir:E69484		E69484	
Description						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
188905_c1_179	218	5440	72	219		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
19645166_f3_136	219	5441	345	1038	244	8.2e-20

Protein name

Locus Name

Acc#

sp:YQEN_BACSU

P54459

Description

HYPOTHETICAL 40.5 KD PROTEIN IN COMEC-RPST INTERGENIC REGION

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
2051500_c3_266	220	5442	777	2334	178	2.5e-23

Protein name

Locus Name

Acc#

conserved hypothetical protein aq_1386

pir:F70420

F70420

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
20741703_f2_89	221	5443	74	225		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
2151556_f1_27	222	5444	99	300		

Protein name Locus Name Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
21541425_c2_199	223	5445	61	186		

Protein name Locus Name Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
21679062_f1_18	224	5446	283	852	206	1.3e-16

Protein name Locus Name Acc#

conserved hypothetical protein pir:E72209 E72209

Description

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
235762_f1_22	225	5447	179	540		

Protein name Locus Name Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
23625693_f2_64	226	5448	265	798		

Protein name Locus Name Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
23633312_f1_7	227	5449	248	747	193	3.1e-15

Protein name

Locus Name

Acc#

gp:APU72238

U72238

Description

Anabaena PCC7120 ORFR1, ORFR2, ORFR3, ORFR4, and ORFR5 genes, complete sequences.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
23635812_f1_9	228	5450	784	2355	148	2.0e-09

Protein name

Locus Name

Acc#

conserved hypothetical protein AF1017

pir:A69377

A69377

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
23884561_f1_45.....	229	5451	131	396	136	4.1e-08

Protein name

Locus Name

Acc#

63 kDa protein

gp:MBU73653

U73653

Description

Mycobacterium bovis 63 kDa protein, 47 kDa protein and clpB gene, complete cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24256502_f3_115.....	230	5452	383	1152		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
24316061_f1_6	231	5453	60	183		

Protein name Locus Name Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
24334393_c1_151	232	5454	98	297		

Protein name Locus Name Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
24412502_c1_196	233	5455	94	285		

Protein name Locus Name Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
24485926_f1_16	234	5456	476	1431		

Protein name Locus Name Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
24494017_c1_169	235	5457	295	888	887	8.9e-89

Protein name Locus Name Acc#

hypothetical protein jhp0694

pir:F71901

F71901

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24500032_f3_121	236	5458	612	1839	1501	7.7e-154

Protein name

Locus Name

Acc#

sp:SYD_BACSU

032038

Description

(ASPRS)

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24642760_c3_311	237	5459	415	1245	1990	1.2e-205

Protein name

Locus Name

Acc#

L-fucose isomerase

gp:AF137263

AF137263

Description

Bacteroides thetaiotaomicron 30S ribosomal protein S16-like protein, fucose gene cluster, and RNA polymerase sigma factor SigZ-like protein (sigZ) genes, complete cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24726550_f2_61	238	5460	136	411		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24804712_f3_114	239	5461	164	495	88	0.013

Protein name

Locus Name

Acc#

ATP synthase F0, subunit b'

pir:A64662

A64662

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24897943_f1_12	240	5462	61	186		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
2531692_f1_21	241	5463	105	318	217	8.9e-18

Protein name

Locus Name

Acc#

gp:AB024563

AB024563

Description

Bacillus halodurans gene for YFIL, YFIM, YFIN, YHDE, HMP and ARGE, complete cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
2617125_f1_40	242	5464	61	186		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
26376540_c1_194	243	5465	255	768	106	0.0015

Protein name

Locus Name

Acc#

sensory transduction system regulatory protein slr1837:protein slr1837:protein slr1837

pir:S77341

S77341

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
26578375_f3_100	244	5466	67	204		

Protein name Locus Name Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
2741543_f3_111	245	5467	608	1827	111	0.0053

Protein name Locus Name Acc#

sp:SECY_ANTSP Q37143

Description

PREPROTEIN TRANSLOCASE SECY SUBUNIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
2851577_f1_41	246	5468	297	894	296	3.8e-26

Protein name Locus Name Acc#

XylR gp:BSU15985 U15985

Description

Bacillus stearothermophilus endo-beta-1,4-xylanase (xynA) gene, complete cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
29344652_f1_37	247	5469	279	840	433	1.1e-40

Protein name Locus Name Acc#

sp:PYZ_BACSU P25983

Description

DIHYDROOROTATE DEHYDROGENASE ELECTRON TRANSFER SUBUNIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
29537532_f3_117	248	5470	101	306	113	9.3e-07
Protein name			Locus Name			Acc#
hypothetical protein Rv2816c			pir:C70691			C70691
Description						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
33205013_c1_157	249	5471	247	744	605	6.8e-59
Protein name			Locus Name			Acc#
			sp:TRMD_BACSU			031741
Description						
METHYLTRANSFERASE)						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
34015677_f3_131	250	5472	279	840	218	7.0e-18
Protein name			Locus Name			Acc#
hypothetical protein CT398			pir:A71519			A71519
Description						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
34068760_f1_28	251	5473	452	1359	599	2.9e-58
Protein name			Locus Name			Acc#
conserved hypothetical protein yqfO			pir:A69954			A69954
Description						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
35345057_c2_240	252	5474	63	192		
Protein name			Locus Name			Acc#
Description						
NO-HIT						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
35432303_c2_228	253	5475	396	1191	778	3.2e-77

Protein name	Locus Name	Acc#
hypothetical protein HP0049	pir:A64526	A64526
Description		

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
35832062_f3_108	254	5476	815	2448	173	2.9e-09

Protein name	Locus Name	Acc#
	sp:YBJZ_ECOLI	P75831
Description		

HYPOTHETICAL ABC TRANSPORTER ATP-BINDING PROTEIN YBJZ

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
36362551_c2_232	255	5477	396	1191	801	1.2e-79

Protein name	Locus Name	Acc#
	sp:BIOF_BACSH	P22806
Description		

LIGASE)

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
3906625_c3_309	256	5478	189	570		

Protein name	Locus Name	Acc#
Description		

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
-----------------	-------------	-------------	------------------	------------------	--------------	--------------------

3923465_f3_135	257	5479	261	786	510	7.9e-49
----------------	-----	------	-----	-----	-----	---------

<u>Protein name</u>	<u>Locus Name</u>	<u>Acc#</u>
---------------------	-------------------	-------------

amp nucleosidase	pir:A72021	A72021
------------------	------------	--------

Description

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
-----------------	-------------	-------------	------------------	------------------	--------------	--------------------

3928336_c2_262.....	258	5480	464	1395		
---------------------	-----	------	-----	------	--	--

<u>Protein name</u>	<u>Locus Name</u>	<u>Acc#</u>
---------------------	-------------------	-------------

Description

NO-HIT		
--------	--	--

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
-----------------	-------------	-------------	------------------	------------------	--------------	--------------------

3932756_f2_74.....	259	5481	471	1416	518	1.1e-49
--------------------	-----	------	-----	------	-----	---------

<u>Protein name</u>	<u>Locus Name</u>	<u>Acc#</u>
---------------------	-------------------	-------------

OprM	gp:AB011381	AB011381
------	-------------	----------

Description

Pseudomonas aeruginosa gene for OprM, complete cds.		
---	--	--

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
-----------------	-------------	-------------	------------------	------------------	--------------	--------------------

3942137_f2_54.....	260	5482	135	408		
--------------------	-----	------	-----	-----	--	--

<u>Protein name</u>	<u>Locus Name</u>	<u>Acc#</u>
---------------------	-------------------	-------------

Description

NO-HIT		
--------	--	--

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
3946962_c1_163	261	5483	63	192		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
3985640_c2_260.....	262	5484	152	459		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
4103383_c2_201.....	263	5485	125	378	164	3.7e-12

Protein name

Locus Name

Acc#

sp:YBDF_ECOLI

Description

HYPOTHETICAL 14.1 KD PROTEIN IN NFNB-ENTD INTERGENIC REGION

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
41307_f1_14.....	264	5486	247	744	187	1.3e-14

Protein name

Locus Name

Acc#

sp:Y978_METJA

Q58388

Description

HYPOTHETICAL PROTEIN MJ0978

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
4343942_f2_75	265	5487	383	1152	443	1.0e-41
<u>Protein name</u>	<u>Locus Name</u>				<u>Acc#</u>	
hypothetical protein mexE	pir:T30829				T30829	
<u>Description</u>						

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
4531536_f3_112.....	266	5488	412	1239	150	7.0e-10
<u>Protein name</u>	<u>Locus Name</u>				<u>Acc#</u>	
	gp:YP102KB				AL031866	
<u>Description</u>						

Yersinia pestis 102 kbases unstable region: from 1 to 119443.

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
4535677_c1_154.....	267	5489	211	636	241	2.5e-20
<u>Protein name</u>	<u>Locus Name</u>				<u>Acc#</u>	
	sp:CAT3_ECOLI				P00484	
<u>Description</u>						

CHLORAMPHENICOL ACETYLTRANSFERASE III,

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
4691032_c2_233.....	268	5490	143	432		
<u>Protein name</u>	<u>Locus Name</u>				<u>Acc#</u>	
<u>Description</u>						

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4691061_c3_294	269	5491	94	285	72	0.020

Protein name	Locus Name	Acc#
NADH dehydrogenase subunit 4L	gp:BMMITOH01	AF110610

Description

Boophilus microplus NADH dehydrogenase subunit 4 (ND4) gene, partial cds; NADH dehydrogenase subunit 4L (ND4L) gene, complete cds; tRNA-Thr and tRNA-Pro genes, complete sequence; and NADH dehydrogenase subunit 6 (ND6) gene, partial cds, mitochondrial genes for mitochondrial products.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4897087_f1_19	270	5492	139	420		

Protein name	Locus Name	Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4898537_f1_11	271	5493	749	2250	136	8.6e-06

Protein name	Locus Name	Acc#
	sp:Y797_METJA	Q58207

Description

HYPOTHETICAL PROTEIN MJ0797

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4973765_f3_104	272	5494	141	426	128	1.6e-07

Protein name	Locus Name	Acc#
conserved hypothetical protein yknZ	pir:E69858	E69858

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
5084381_c3_310	273	5495	331	996	1166	2.4e-118

Protein name

Locus Name

Acc#

FucR

gp:AF137263

AF137263

Description

Bacteroides thetaiotaomicron 30S ribosomal protein S16-like protein, fucose gene cluster, and RNA polymerase sigma factor SigZ-like protein (sigZ) genes, complete cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
5187812_f1_13	274	5496	228	687	596	6.1e-58

Protein name

Locus Name

Acc#

sp:YF08_METJA

Q58903

Description

HYPOTHETICAL ABC TRANSPORTER ATP-BINDING PROTEIN MJ1508

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
5191942_f3_128	275	5497	209	630	246	7.5e-21

Protein name

Locus Name

Acc#

amino acid ABC transporter, ATP-binding protein

pir:H72356

H72356

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
547082_f1_36	276	5498	170	513	93	0.00016

Protein name

Locus Name

Acc#

gpC

gp:AF063097

Description

Bacteriophage P2, complete genome.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
5859625_c3_299	277	5499	62	189		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
6412812_c2_198	278	5500	686	2061	2246	8.7e-233

Protein name

Locus Name

Acc#

high temperature protein HtpG

gp:AF176245

AF176245

Description

Porphyromonas gingivalis high temperature protein HtpG (htpG) gene, complete cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
6723262_f2_56	279	5501	133	402		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
787567_f3_139	280	5502	303	912	659	1.3e-64

Protein name

Locus Name

Acc#

dihydrodipicolinate synthase

pir:B72246

B72246

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
907974_c1_149	281	5503	745	2238	1877	1.1e-193

Protein name

Locus Name

Acc#

sp:MECB_BACSU

P37571

Description

NEGATIVE REGULATOR OF GENETIC COMPETENCE MECB

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
9851505_f3_113	282	5504	238	717	106	1.5e-05

Protein name

Locus Name

Acc#

hypothetical protein

gp:SEL243707

AJ243707

Description

Synechococcus elongatus petB gene, petD gene and ORF1.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
10969062_f3_151.....	283	5505	256	771	747	6.1e-74

Protein name

Locus Name

Acc#

ATP synthase F1, subunit alpha

pir:F72231

F72231

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
133562_f3_116.....	284	5506	265	798	565	1.2e-54

Protein name

Locus Name

Acc#

hypothetical protein

gp:STE242827

AJ242827

Description

Streptomyces tendae atp gene and ORF2 (partial), strain Tue901/8c.

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
13672255_f2_95	285	5507	146	441		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
14501312_c3_311.....	286	5508	461	1386	473	6.6e-45

Protein name

Locus Name

Acc#

conserved hypothetical integral membrane protein HP1184

pir:H64667

H64667

Description

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
14511007_f1_2.....	287	5509	330	993		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
14645437_f3_135.....	288	5510	768	2307	117	0.0015

Protein name

Locus Name

Acc#

conserved hypothetical protein yknZ

pir:E69858

E69858

Description

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
14969692_f1_37.....	289	5511	404	1215	232	4.9e-17

Protein name

Locus Name

Acc#

antibiotic resistance protein homolog ywoG

pir:B70065

B70065

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
15104137_f1_1	290	5512	466	1401		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
15673443_c3_323	291	5513	464	1395	792	1.0e-78

Protein name

Locus Name

Acc#

Salmonella typhimurium transcriptional

gp:STYSTMF1

AF170176

Description

Salmonella typhimurium fragment STMF1.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
158187_f3_121	292	5514	409	1230	988	1.8e-99

Protein name

Locus Name

Acc#

sp:URAA_HAEIN

P45117

Description

PROBABLE URACIL PERMEASE (URACIL TRANSPORTER)

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
16172682_f1_45	293	5515	310	933	532	3.4e-78

Protein name

Locus Name

Acc#

sp:ATPA_RICPR

O50288

Description

ATP SYNTHASE ALPHA CHAIN,

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
16828462_f1_46	294	5516	292	879	470	1.4e-44
<u>Protein name</u>			<u>Locus Name</u>		<u>Acc#</u>	
			sp:ATPG_BACSU		P37810	

Description

ATP SYNTHASE GAMMA CHAIN,

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
19945317_c2_264	295	5517	285	858	120	2.0e-05
<u>Protein name</u>			<u>Locus Name</u>		<u>Acc#</u>	
3',5'-cyclic-nucleotide phosphodiesterase, cpdA homolog MTH178:Icc related protein			pir:F69104		F69104	

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
21990930_c1_217.....	296	5518	326	981		
<u>Protein name</u>			<u>Locus Name</u>		<u>Acc#</u>	

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
22464182_c1_215.....	297	5519	328	987		
<u>Protein name</u>			<u>Locus Name</u>		<u>Acc#</u>	

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
2346936_f3_153	298	5520	463	1392	153	8.5e-08

<u>Protein name</u>	<u>Locus Name</u>	<u>Acc#</u>
HeIC	gp:LPU11704	U11704

Description

Legionella pneumophila HeIC (heIC) gene, complete cds.

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
23694686_f1_35	299	5521	225	678	179	9.5e-14

<u>Protein name</u>	<u>Locus Name</u>	<u>Acc#</u>
	sp:GS1_HUMAN	Q08623

Description

GS1 PROTEIN

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
23829510_c1_325.....	300	5522	228	687	283	9.0e-25

<u>Protein name</u>	<u>Locus Name</u>	<u>Acc#</u>
transcription regulator, crp family	pir:F72285	F72285

Description

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
24009530_c1_182.....	301	5523	70	213		

<u>Protein name</u>	<u>Locus Name</u>	<u>Acc#</u>

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24010962_f1_25	302	5524	131	396		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24111375_f3_110	303	5525	68	207		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24219032_f2_97	304	5526	92	279	172	5.2e-13

Protein name

Locus Name

Acc#

sp:ATPL_ANASP

P12409

Description

ATP SYNTHASE C CHAIN, (LIPID-BINDING PROTEIN)

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24254417_f2_53	305	5527	379	1140	480	1.2e-45

Protein name

Locus Name

Acc#

sensory transduction system regulatory
protein sll1229:protein sll1229:protein
sll1229

pir:S75524

S75524

Description

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
24333127_c3_332	306	5528	532	1599	748	1.3e-76

Protein name

Locus Name

Acc#

sp:Y1EN_ECOLI

P31473

Description

HYPOTHETICAL 56.4 KD PROTEIN IN ASNA-KUP INTERGENIC REGION

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
24398417_c2_280	307	5529	996	2991	108	3.4e-15

Protein name

Locus Name

Acc#

hypothetical protein jhp0336

pir:C71944

C71944

Description

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
24399035_c3_312	308	5530	60	183	44	0.049

Protein name

Locus Name

Acc#

nonstructural protein

gp:AF012732

AF012732

Description

Bovine viral diarrhea virus strain Yak nonstructural protein (p125)mRNA, partial cds.

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
24408437_c2_228	309	5531	122	369	265	7.3e-23

Protein name

Locus Name

Acc#

sp:THIO_BORBU

O51088

Description

THIOREDOXIN (TRX)

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24414153_f1_20	310	5532	448	1347	83	0.0040

Protein name	Locus Name	Acc#
unknown	gp:U96771	U96771

Description

Prevotella bryantii putative polygalacturonase, B-1,4-endoglucanase, and mannanase genes, complete cds; and unknown genes.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24489452_c2_269	311	5533	560	1683	526	2.0e-58

Protein name	Locus Name	Acc#
long-chain-fatty-acid CoA ligase	pir:D70386	D70386

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24644068_f2_94	312	5534	83	252	158	1.6e-11

Protein name	Locus Name	Acc#
	sp:ATPE_CHLLI	P35111

Description

ATP SYNTHASE EPSILON CHAIN,

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24808312_f1_49	313	5535	189	567	308	3.8e-26

Protein name	Locus Name	Acc#
	sp:HELA_LEGPN	Q48815

Description

HELA PROTEIN

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24875042_f3_147	314	5536	76	231		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24881577_f2_85	315	5537	248	747	210	5.1e-16

Protein name

Locus Name

Acc#

sp:XYNB_BUTFI

P26223

Description

D-XYLAN XYLANOHYDROLASE B)

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
25479842_c3_315	316	5538	67	204		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
26446928_f3_113	317	5539	254	765	129	3.7e-05

Protein name

Locus Name

Acc#

receptor antigen (RagA)

gp:PGI130872

AJ130872

Description

Porphyromonas gingivalis W50 receptor antigen (rag) locus encoding a major immunodominant 55kDa antigen.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
26831386_f3_139	318	5540	349	1050	507	1.7e-48

Protein name

Locus Name

Acc#

sp:PYRD_ECOLI

P05021

Description

(DHODHASE)

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
2942827_c3_342	319	5541	1057	3174	214	3.5e-28

Protein name

Locus Name

Acc#

probable ATP-dependent helicase

pir:A71805

A71805

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
29494000_f1_48	320	5542	364	1095	309	4.8e-27

Protein name

Locus Name

Acc#

sp:CZCB_ALCSP

P94176

Description

CATION EFFLUX SYSTEM PROTEIN CZCB

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
29572678_f2_63	321	5543	208	627	213	5.3e-16

Protein name

Locus Name

Acc#

115K outer membrane protein precursor:SusC protein

pir:JC6027

JC6027

Description

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT</u> <u>Length</u>	<u>AA</u> <u>Length</u>	<u>Score</u>	<u>Probability</u>
29973182_c2_246	322	5544	857	2574	1042	3.4e-105

Protein name

Locus Name

Acc#

(p)ppGpp synthetase

gp:BSU86377

U86377

Description

Bacillus subtilis (p)ppGpp synthetase (relA) and adeninephosphoribosyltransferase (apt) genes, complete cds.

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT</u> <u>Length</u>	<u>AA</u> <u>Length</u>	<u>Score</u>	<u>Probability</u>
3131910_c2_268	323	5545	492	1479	832	6.0e-83

Protein name

Locus Name

Acc#

sp:YCGO_ECOLI

P76007

Description

PUTATIVE NA(+)/H(+) EXCHANGER YCGO

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT</u> <u>Length</u>	<u>AA</u> <u>Length</u>	<u>Score</u>	<u>Probability</u>
3165800_c3_302.....	324	5546	103	312	134	5.5e-09

Protein name

Locus Name

Acc#

ORF2

gp:AB015879

AB015879

Description

Porphyromonas gingivalis dnaK operon genes, complete cds.

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT</u> <u>Length</u>	<u>AA</u> <u>Length</u>	<u>Score</u>	<u>Probability</u>
32660011_f3_150.....	325	5547	415	1248	276	5.8e-24

Protein name

Locus Name

Acc#

sp:ATP6_RHORU

P15012

Description

ATP SYNTHASE A CHAIN, (PROTEIN 6)

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
33240828_f2_62	326	5548	885	2658		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
33788387_f2_78	327	5549	310	933	101	0.024

Protein name

Locus Name

Acc#

conserved hypothetical protein

pir:G72385

G72385

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
33992211_c1_181	328	5550	132	399	200	5.6e-16

Protein name

Locus Name

Acc#

diacylglycerol kinase

gp:BSU29177

U29177

Description

Bacillus subtilis PhoH (phoH) gene, partial cds, diacylglycerolkinase (dgg) gene, complete cds, and cytidine deaminase (cdd) gene, partial cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
34178385_f3_149	329	5551	511	1536	2540	6.1e-264

Protein name

Locus Name

Acc#

sp:ATPB_BACFR

P13356

Description

ATP SYNTHASE BETA CHAIN,

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
34181561_f1_19	330	5552	602	1809	452	8.9e-52
Protein name			Locus Name		Acc#	
115K outer membrane protein precursor:SusC protein			pir:JC6027		JC6027	
Description						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
35625000_f2_92	331	5553	530	1593		
Protein name			Locus Name		Acc#	
Description			NO-HIT			

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
3937751_f2_101	332	5554	864	2595	404	1.3e-36
Protein name			Locus Name		Acc#	
DNA helicase homolog			gp:AF108138		AF108138	
Description						
Homo sapiens DNA helicase homolog (PIF1) mRNA, partial cds.						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4395252_c1_209	333	5555	528	1587	494	1.7e-62
Protein name			Locus Name		Acc#	
Beta-N-Acetylglucosaminidase			gp:AB015350		AB015350	
Description						
Streptomyces thermoviolaceus nagB gene forBeta-N-Acetylglucosaminidase, complete cds.						

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
4454637_f1_5	334	5556	315	948	1058	6.8e-107

Protein name

Locus Name

Acc#

dTDP-glucose 4-6-dehydratase:protein
slr0809:protein slr0809

pir:S75550

S75550

Description

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
4501875_f2_51	335	5557	124	375		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
4553376_c1_199	336	5558	459	1380	291	2.5e-25

Protein name

Locus Name

Acc#

gp:ECOUW82

L10328

Description

E. coli; the region from 81.5 to 84.5 minutes.

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
4881660_c2_266	337	5559	547	1644	1809	1.8e-186

Protein name

Locus Name

Acc#

sp:PRIS_DESVH

P31101

Description

PRISMANE PROTEIN

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
5276557_f2_98	338	5560	169	510	197	1.2e-15
<u>Protein name</u>			<u>Locus Name</u>		<u>Acc#</u>	
ATP synthase F0, subunit b			pir:H72231		H72231	
<u>Description</u>						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
5276586_f2_99	339	5561	189	570	215	1.4e-17
<u>Protein name</u>			<u>Locus Name</u>		<u>Acc#</u>	
F1F0-ATPase subunit delta			gp:AF098522		AF098522	
<u>Description</u>						
Lactobacillus acidophilus uracil phosphoribosyltransferase (upp) gene, partial cds; and F1F0-ATPase operon, complete sequence.						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
556557_f3_134	340	5562	478	1437	136	6.4e-06
<u>Protein name</u>			<u>Locus Name</u>		<u>Acc#</u>	
			sp:YF07_METJA		Q58902	
<u>Description</u>						
HYPOTHETICAL PROTEIN MJ1507						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
5890667_c1_176	341	5563	95	288	213	2.4e-17
<u>Protein name</u>			<u>Locus Name</u>		<u>Acc#</u>	
RNA-binding protein			gp:ANARBPD2		D49425	
<u>Description</u>						
Anabaena variabilis rbpD gene for RNA-binding protein, complete cds.						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
6046881_f3_140	342	5564	306	921	655	3.4e-64
<u>Protein name</u>			<u>Locus Name</u>		<u>Acc#</u>	
3-methyl-2-oxobutanoate			gp:CGPAN		X96580	
<u>Description</u>						
C.glutamicum panB, panC & xylB genes.						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
6836013_c2_265	343	5565	443	1332	281	1.1e-21
<u>Protein name</u>			<u>Locus Name</u>		<u>Acc#</u>	
			sp:NTRY_AZOCA		Q04850	
<u>Description</u>						
NITROGEN REGULATION PROTEIN NTRY,						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
7072037_c3_322	344	5566	240	723		
<u>Protein name</u>			<u>Locus Name</u>		<u>Acc#</u>	
<u>Description</u>						
NO-HIT						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
814126_f2_60	345	5567	84	255		
<u>Protein name</u>			<u>Locus Name</u>		<u>Acc#</u>	
<u>Description</u>						
NO-HIT						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
9766325_f3_148	346	5568	422	1269	1070	3.6e-108

Protein name

Locus Name

Acc#

sp:PUPT_SYNY3

Q55336

Description

2) (FORMATE-DEPENDENT GAR TRANSFORMYLASE)

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
1057762_f2_175	347	5569	214	645	528	9.8e-51

Protein name

Locus Name

Acc#

thio-specific antioxidant (tsa) peroxidase

pir:E72036

E72036

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
1173263_c2_422	348	5570	77	234		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
11834382_c2_479	349	5571	299	900	79	0.021

Protein name

Locus Name

Acc#

ATP binding protein

gp:BBATPBP

X91965

Description

B.burgdorferi abp gene.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
118906_c1_393	350	5572	165	498		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
12204402_c2_483	351	5573	67	204	72	0.020

Protein name

Locus Name

Acc#

pE66L

gp:ASU18466

U18466

Description

African swine fever virus, complete genome.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
12677202_t3_307	352	5574	237	714	472	8.5e-45

Protein name

Locus Name

Acc#

hypothetical protein

gp:AHAAMYG

X58627

Description

A.haloplanktis amy gene for alpha-amylase
1,4-alpha-D-glucanglucanohydrolase.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
13103938_c2_466	353	5575	159	480	227	5.6e-22

Protein name

Locus Name

Acc#

single stranded DNA-binding protein

gp:SSU64095

U64095

Description

Shewanella sp. PT99 single stranded DNA-binding protein (ssb) gene, complete cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
1367792_f3_258	354	5576	202	609	298	2.3e-26

Protein name	Locus Name	Acc#
	sp:YB69_HAEIN	P44118

Description

HYPOTHETICAL PROTEIN HI1169

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
1369082_c1_320	355	5577	419	1260	714	3.8e-75

Protein name	Locus Name	Acc#
autoaggregation-mediating protein	gp:AF091502	AF091502

Description

Lactobacillus reuteri autoaggregation-mediating protein (aggH) gene, complete cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
13787827_c2_468.....	356	5578	156	471		

Protein name	Locus Name	Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
13876713_f2_124.....	357	5579	103	312		

Protein name	Locus Name	Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
13914808_f2_181	358	5580	103	312		

Protein name Locus Name Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
14304550_c2_490	359	5581	380	1143		

Protein name Locus Name Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
1441937_c1_386	360	5582	322	969		

Protein name Locus Name Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
14494037_c3_524	361	5583	262	789	256	1.5e-32

Protein name Locus Name Acc#

sp:CCSA_CYACA P31564

Description

CYTOCHROME C BIOGENESIS PROTEIN CCSA

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
14660927_c1_394	362	5584	1249	3750	259	7.8e-18

Protein name

Locus Name

Acc#

gp:SCYDL057W

Description

S.cerevisiae chromosome IV reading frame ORF YDL057w.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
14665882_c2_461	363	5585	153	462		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
14875191_c1_388.....	364	5586	365	1098	116	0.00061

Protein name

Locus Name

Acc#

hypothetical protein

gp:YEN132945

AJ132945

Description

Yersinia enterocolitica WA 314 right arm of the high-pathogenicityisland.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
156325_f2_176.....	365	5587	116	351	211	3.8e-17

Protein name

Locus Name

Acc#

ss-DNA binding protein 12RNP2 precursor

gp:SYO12RNP2

D17359

Description

Synechococcus 6301 gene for ss-DNA binding protein 12RNP2, completecds.

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
15660937_c1_345	366	5588	432	1299	617	3.6e-60
<u>Protein name</u>			<u>Locus Name</u>		<u>Acc#</u>	
hypothetical protein			pir:T33724		T33724	
<u>Description</u>						

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
162785_c3_565	367	5589	322	969	156	7.4e-09
<u>Protein name</u>			<u>Locus Name</u>		<u>Acc#</u>	
Mag44			gp:DEPMAG44		D17682	
<u>Description</u>						

Dermatophagoides farinae mRNA for Mag44, partial cds.

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
16447875_c3_560	368	5590	63	192		
<u>Protein name</u>			<u>Locus Name</u>		<u>Acc#</u>	
<u>Description</u>						

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
16525318_c2_471	369	5591	574	1725		
<u>Protein name</u>			<u>Locus Name</u>		<u>Acc#</u>	
<u>Description</u>						

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
165882_c1_371	370	5592	1018	3057	179	7.8e-10
<u>Protein name</u>			<u>Locus Name</u>		<u>Acc#</u>	
			sp:PRIM_CLOAB		P33655	
<u>Description</u>						
DNA PRIMASE,						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
16900087_c2_420	371	5593	279	840	414	1.2e-38
<u>Protein name</u>			<u>Locus Name</u>		<u>Acc#</u>	
hypothetical protein yycJ			pir:A70090		A70090	
<u>Description</u>						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
17010890_c1_369	372	5594	113	342		
<u>Protein name</u>			<u>Locus Name</u>		<u>Acc#</u>	
<u>Description</u>						
NO-HIT						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
180302_c3_559	373	5595	73	222		
<u>Protein name</u>			<u>Locus Name</u>		<u>Acc#</u>	
<u>Description</u>						
NO-HIT						

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
187930_c3_509	374	5596	359	1080		

Protein name Locus Name Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
19035_f1_47	375	5597	674	2025	1563	2.1e-160

Protein name Locus Name Acc#

branching enzyme gp:AB026630 AB026630

Description

Emericella nidulans gene for branching enzyme, complete cds.

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
197191_f3_291	376	5598	381	1146	129	3.1e-05

Protein name Locus Name Acc#

sp:PORP_PSEAE P05695

Description

PORIN P PRECURSOR (OUTER MEMBRANE PROTEIN D1)

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
19729591_c1_380	377	5599	171	516		

Protein name Locus Name Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
19734688_c2_423	378	5600	193	582	96	0.021

Protein name

Locus Name

Acc#

two-component sensor histidine kinase homolog
ybdK

pir:F69747

F69747

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
19740893_c3_542	379	5601	408	1227	1169	1.2e-118

Protein name

Locus Name

Acc#

sp:FSR_ECOLI

P52067

Description

FOSMIDOMYCIN RESISTANCE PROTEIN

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
1988388_f1_44	380	5602	64	195		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
20314007_c2_407	381	5603	827	2484	1786	4.8e-184

Protein name

Locus Name

Acc#

sp:LON1_MYXXA

P36773

Description

ATP-DEPENDENT PROTEASE LA 1,

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
20344086_f2_157	382	5604	65	198		

Protein name _____ Locus Name _____ Acc# _____

Description _____

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
20350260_c2_405	383	5605	70	213		

Protein name _____ Locus Name _____ Acc# _____

Description _____

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
204437_f3_277	384	5606	231	696		

Protein name _____ Locus Name _____ Acc# _____

Description _____

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
20501551_c1_325	385	5607	183	552		

Protein name _____ Locus Name _____ Acc# _____

Description _____

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
20980313_c1_392	386	5608	69	210	47	0.034

Protein name

Locus Name

Acc#

sp:YOR5_TTV1

P19280

Description

HYPOTHETICAL 9.5 KD PROTEIN

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
2126506_c1_314	387	5609	170	513	93	0.00016

Protein name

Locus Name

Acc#

transcription regulator phage-related homolog ydcN

pir:C69774

C69774

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
21485027_c2_481.....	388	5610	192	579		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
21490925_c1_365.....	389	5611	110	333		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
21664055_c3_585	390	5612	81	246	69	0.042

Protein name	Locus Name	Acc#
ATP synthase gamma chain	gp:AB027877	AB027877

Description

Schizosaccharomyces pombe gene for ATP synthase gamma chain,partial cds, clone:TA25.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
21677180_c3_566	391	5613	106	321		

Protein name	Locus Name	Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
21688925_f1_40.....	392	5614	66	201	56	0.031

Protein name	Locus Name	Acc#
estrogen receptor	pir:S26595	S26595

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
21729812_c2_408.....	393	5615	374	1125	236	1.0e-17

Protein name	Locus Name	Acc#
hypothetical protein slr0882	pir:S77272	S77272

Description

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
21756268_c2_419	394	5616	81	246		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
22057162_f1_53	395	5617	506	1521	388	7.9e-41

Protein name

Locus Name

Acc#

sp:GLNA_BACCE

P19064

Description

GLUTAMINE SYNTHETASE, (GLUTAMATE--AMMONIA LIGASE)

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
22087762_c3_496	396	5618	189	570		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
22132811_f2_180	397	5619	76	231		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
22459802_f2_159	398	5620	333	1002	759	3.3e-75
<u>Protein name</u>			<u>Locus Name</u>		<u>Acc#</u>	
p-aminobenzoate synthase component I homolog			pir:F64187		F64187	
<u>Description</u>						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
22462807_c2_476	399	5621	178	537	93	0.042
<u>Protein name</u>			<u>Locus Name</u>		<u>Acc#</u>	
			sp:TGN3_RAT		P19814	
<u>Description</u>						

TRANS-GOLGI NETWORK INTEGRAL MEMBRANE PROTEIN TGN38 PRECURSOR

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
22656553_f3_296	400	5622	275	828		
<u>Protein name</u>			<u>Locus Name</u>		<u>Acc#</u>	
<u>Description</u>						

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
23438887_c1_389	401	5623	126	381	87	0.013
<u>Protein name</u>			<u>Locus Name</u>		<u>Acc#</u>	
unknown			gp:AF074396		AF074396	
<u>Description</u>						

Desulfotomaculum thermocisternum
 UDP-acetylglucosamine1-carboxyvinyltransferase (murA) gene, partial cds;
 yydA,ferredoxin (fdx), dissimilatory sulfite reductase subunit A
 (dsrA),dissimilatory sulfite reductase subunit B (dsrB), and dsrD
 genes,complete cds; and unknown gene.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
23472178_f3_271	402	5624	470	1413	1205	1.8e-122

Protein name

Locus Name

Acc#

Xylose Isomerase

gp:RFL132472

AJ132472

Description

Ruminococcus flavefaciens xylan utilization operon.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
23597202_c3_513	403	5625	198	597	48	0.039

Protein name

Locus Name

Acc#

hypothetical protein F21D9.3

pir:T21205

T21205

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
23632132_f1_64	404	5626	510	1533	614	7.6e-60

Protein name

Locus Name

Acc#

xylulose kinase

gp:AF001974

AF001974

Description

Thermoanaerobacter ethanolicus putative TrkG gene, partial cds, and putative TrkA, xylose isomerase (xylA) and xylulose kinase (xylB) genes, complete cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
23634708_c1_368	405	5627	84	255	69	0.042

Protein name

Locus Name

Acc#

sp:YC13_METJA

Q58610

Description

HYPOTHETICAL PROTEIN MJ1213

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
23651680_f2_191	406	5628	78	237		

Protein name _____ Locus Name _____ Acc# _____

Description _____

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
23690875_c1_318	407	5629	385	1158	931	1.9e-93

Protein name _____ Locus Name _____ Acc# _____

sp:TGT_BACSU 032053

Description _____

TRANSGLYCOSYLASE) (GUANINE INSERTION ENZYME)

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
23710926_f2_184	408	5630	355	1068	82	0.013

Protein name _____ Locus Name _____ Acc# _____

M protein precursor

pir:S61081

Description _____

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
23865651_c1_327	409	5631	60	183		

Protein name _____ Locus Name _____ Acc# _____

Description _____

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
23922135_c3_495	410	5632	133	402	73	0.016

<u>Protein name</u>	<u>Locus Name</u>	<u>Acc#</u>
MesF	gp:AF143443	AF143443

Description

Leuconostoc mesenteroides plasmid pHY30 MesG (mesG) gene, partialcds; and mesentericin B105 (mesB), MesH (mesH), and MesF (mesF) genes, complete cds.

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
24027213_c2_460	411	5633	206	621		

<u>Protein name</u>	<u>Locus Name</u>	<u>Acc#</u>

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
24241261_f1_77	412	5634	229	690		

<u>Protein name</u>	<u>Locus Name</u>	<u>Acc#</u>

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
24265886_c2_469	413	5635	158	477		

<u>Protein name</u>	<u>Locus Name</u>	<u>Acc#</u>

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
24305437_c3_556	414	5636	80	243		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
24322127_c3_567.....	415	5637	202	609		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
24344641_f3_233.....	416	5638	120	363		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
24397877_c1_374.....	417	5639	64	195	71	0.0075

Protein name

Locus Name

Acc#

sp:RAFR_ECOLI

P21867

Description

RAFFINOSE OPERON REPRESSOR

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24406557_f2_158	418	5640	252	759	151	4.2e-08

Protein name	Locus Name	Acc#
protein antigen LmSTII	gp:LMU73845	U73845

Description

Leishmania major protein antigen LmSTII mRNA, partial cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24415912_f3_257	419	5641	122	369	80	0.0029

Protein name	Locus Name	Acc#
putative repressor protein	gp:BA1242593	AJ242593

Description

Bacteriophage A118 complete genome.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24484682_c2_492.....	420	5642	359	1077		

Protein name	Locus Name	Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24503282_c3_538.....	421	5643	83	252	78	0.017

Protein name	Locus Name	Acc#
hypothetical protein MJ1664	pir:F64507	F64507

Description

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
24633387_c1_354	422	5644	331	996	635	4.5e-62
<u>Protein name</u>			<u>Locus Name</u>		<u>Acc#</u>	
hypothetical protein T27E13.6			pir:T00580		T00580	
<u>Description</u>						

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
24640915_c2_480.....	423	5645	157	474		
<u>Protein name</u>			<u>Locus Name</u>		<u>Acc#</u>	
<u>Description</u>						
NO-HIT						

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
24644636_c1_372.....	424	5646	111	336		
<u>Protein name</u>			<u>Locus Name</u>		<u>Acc#</u>	
<u>Description</u>						
NO-HIT						

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
24645337_f3_295.....	425	5647	439	1320	1245	1.0e-126
<u>Protein name</u>			<u>Locus Name</u>		<u>Acc#</u>	
putative UDP-glucose dehydrogenase			gp:AF159428		AF159428	
<u>Description</u>						

Burkholderia pseudomallei putative UDP-glucose dehydrogenase (udg), putative ADP-heptose synthase (waaE), and putative ADP-glycero-mannoheptose epimerase (gmhD) genes, complete cds.

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
24648412_f1_23	426	5648	166	501		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
24648562_c3_555.....	427	5649	82	249		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
24710926_f3_299.....	428	5650	174	525	121	1.8e-07

Protein name

Locus Name

Acc#

thiol:disulfide interchange protein homolog
yneN

pir:E69891

E69891

Description

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
24782635_f1_93.....	429	5651	187	564	544	2.0e-52

Protein name

Locus Name

Acc#

dTDP-6-deoxy-D-glucose-3,5 epimerase

gp:AF048749

AF048749

Description

Bacteroides fragilis capsular polysaccharide biosynthesis operon, complete sequence.

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
24798568_f2_219	430	5652	206	621		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
24804681_c1_370	431	5653	74	225		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
24804807_f2_160	432	5654	228	687		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
24820925_f1_80	433	5655	113	342		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24882942_c3_544	434	5656	357	1074	666	2.3e-65

Protein name

Locus Name

Acc#

sp:YVAA_BACSU

032223

Description

HYPOTHETICAL OXIDOREDUCTASE IN FHUD-OPUBD INTERGENIC REGION

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
2537802_c2_404	435	5657	181	546	52	0.044

Protein name

Locus Name

Acc#

envelope glycoprotein

gp:AF021739

AF021739

Description

HIV-1 isolate sling clone 45 from the Netherlands, envelope glycoprotein (env) gene, partial cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
25428312...F3...285.....	436	5658	174	525		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
25556532_c2_402.....	437	5659	145	438		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
25572212_c1_315	438	5660	122	369	69	0.042

<u>Protein name</u>	<u>Locus Name</u>	<u>Acc#</u>
hypothetical protein yopO	pir:T12849	
<u>Description</u>		

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
25664086_c1_390	439	5661	303	912		

<u>Protein name</u>	<u>Locus Name</u>	<u>Acc#</u>
<u>Description</u>		

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
2595058_f2_127	440	5662	70	213		

<u>Protein name</u>	<u>Locus Name</u>	<u>Acc#</u>
<u>Description</u>		

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
26433216_c3_499	441	5663	81	246		

<u>Protein name</u>	<u>Locus Name</u>	<u>Acc#</u>
<u>Description</u>		

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
26517_c1_346	442	5664	69	210		

<u>Protein name</u>	<u>Locus Name</u>	<u>Acc#</u>
<u>Description</u>		

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
26601510_c2_448	443	5665	71	216	101	1.7e-05

Protein name	Locus Name	Acc#
hypothetical protein MJ1608	pir:G64500	G64500

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
26601577_c1_359	444	5666	268	807	425	8.1e-40

Protein name	Locus Name	Acc#
conserved hypothetical protein aq_1386	pir:F70420	F70420

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
26692342_c2_446	445	5667	406	1221	856	1.7e-85

Protein name	Locus Name	Acc#
succinate--CoA ligase (ADP-forming), beta chain	pir:H70439	H70439

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
273261_c2_477	446	5668	480	1443		

Protein name	Locus Name	Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
2770305_c3_580	447	5669	214	645		

Protein name	Locus Name	Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
29412901_f1_79	452	5674	221	666	135	7.5e-14
Protein name			Locus Name		Acc#	
			sp:LSPA_STACA		Q59835	
Description						

PEPTIDASE) (SIGNAL PEPTIDASE II) (SPASE II)

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
29470081_c1_376	453	5675	334	1005	102	0.0029
Protein name			Locus Name		Acc#	
hypothetical protein PH0283			pir:D71453		D71453	
Description						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
30289001_c3_525.....	454	5676	468	1407	1084	1.2e-109
Protein name			Locus Name		Acc#	
cytochrome c peroxidase			gp:AF200362		AF200362	
Description						

Haemophilus ducreyi oxaloacetate decarboxylase gamma chain (oadG) gene, partial cds; oxaloacetate decarboxylase alpha chain (oadA), oxaloacetate decarboxylase beta chain (oadB), and alkylphosphonate uptake protein (phna) genes, complete cds; ccp gene, complete sequence; cytochrome c peroxidase gene, complete cds; and unknown gene.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
30588453_f1_1.....	455	5677	63	192		
Protein name			Locus Name		Acc#	
Description						

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
31672502_f1_24	456	5678	262	789	118	0.00032

Protein name	Locus Name	Acc#
type I restriction enzyme hsdM:hypothetical protein H91_orf543:hypothetical protein H91_orf543	pir:S73820	S73820

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
32035967_c3_507	457	5679	121	366	161	7.6e-12

Protein name	Locus Name	Acc#
hypothetical protein	gp:SSU18930	Y18930

Description

Sulfolobus solfataricus 281 kb genomic DNA fragment, strain P2.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
33364762_c1_356.....	458	5680	300	903	890	4.3e-89

Protein name	Locus Name	Acc#
succinate--CoA ligase (ADP-forming), alpha chain	pir:F69719	F69719

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
3385955_c1_360.....	459	5681	586	1761	148	8.5e-16

Protein name	Locus Name	Acc#
hypothetical protein TM1650	pir:G72227	G72227

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
34017140_c3_498	460	5682	64	195		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
34073250_f1_54	461	5683	258	777	234	1.4e-19

Protein name

Locus Name

Acc#

sp:YT29_MYCTU

P71564

Description

PUTATIVE OXIDOREDUCTASE RV0945,

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
34176462_c2_418	462	5684	472	1419	1369	7.5e-140

Protein name

Locus Name

Acc#

sp:UXAC_ECOLI

Description

ISOMERASE)

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
34277280_c1_378	463	5685	257	774		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
34406502_c2_403	464	5686	130	393		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
34407787_f1_55	465	5687	198	597	132	7.8e-07

Protein name

Locus Name

Acc#

sp:Y374_METJA

Q57819

Description

HYPOTHETICAL PROTEIN MJ0374

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
35634587_f1_74	466	5688	532	1599	1249	3.9e-127

Protein name

Locus Name

Acc#

sp:YHCX_BACSU

P54608

Description

HYPOTHETICAL 60.2 KD PROTEIN IN CSPB-GLPP INTERGENIC REGION

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
36337562_c3_553	467	5689	139	420	91	0.00020

Protein name

Locus Name

Acc#

regulatory protein CsgD

gp:ECOCURLI2

AF081826

Description

Escherichia coli csg cluster, partial sequence.

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
36500003_c1_328	468	5690	430	1293	849	9.5e-85

Protein name

Locus Name

Acc#

macrolide-efflux determinant

gp:SPU83667

U83667

Description

Streptococcus pneumoniae macrolide-efflux determinant (mefE) gene, complete cds.

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
3938838_c1_373	469	5691	131	396		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
4021882_c3_563	470	5692	431	1296		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
4062628_c3_577	471	5693	115	348		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4065760_f1_63	472	5694	255	768	239	4.1e-20
Protein name			Locus Name		Acc#	
hypothetical protein			pir:S75926		S75926	
Description						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4072687_f1_81	473	5695	773	2322	121	1.5e-05
Protein name			Locus Name		Acc#	
outer membrane protein			gp:NGU81959		U81959	
Description						
Neisseria gonorrhoeae outer membrane protein (omp85) gene, complete cds.						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4157536_c3_576	474	5696	88	267	77	0.018
Protein name			Locus Name		Acc#	
hypothetical protein ZC47.1			pir:T27592		T27592	
Description						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4318885_f2_174	475	5697	509	1530	1371	4.6e-140
Protein name			Locus Name		Acc#	
xylose transporter			gp:AB009593		AB009593	
Description						
Tetragenococcus halophilus rbsC, rbsB, xylR, xylA, xylB and xylE genes, partial and complete cds.						

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
4345012_f2_166	476	5698	196	591	92	0.028

Protein name Locus Name Acc#

sp:CRP_ECOLI P03020

Description

PROTEIN)

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
4553288_f1_45	477	5699	67	204		

Protein name Locus Name Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
4695285_c2_434.....	478	5700	319	960		

Protein name Locus Name Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
4773400_c3_557.....	479	5701	119	360		

Protein name Locus Name Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
4791400_f2_134	480	5702	71	216		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
4877135_f3_279	481	5703	1149	3450	991	1.7e-211

Protein name

Locus Name

Acc#

isoleucine--tRNA ligase, ileS:isoleucyl-tRNA synthetase:isoleucyl-tRNA synthetase

pir:H70203

H70203

Description

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
4883592_c3_574	482	5704	181	546		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
4978302_c3_504	483	5705	424	1275	659	1.3e-64

Protein name

Locus Name

Acc#

probable phosphoserine phosphatase

pir:T36772

T36772

Description

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
564077_c2_447	484	5706	78	237		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
6022037_c1_337	485	5707	353	1062	764	9.6e-76

Protein name

Locus Name

Acc#

sp:YHIM_ECOLI

Description

HYPOTHETICAL 39.2 KD PROTEIN IN RHSB-PIT INTERGENIC REGION

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
6347188_c3_588	486	5708	1643	4932	161	1.3e-08

Protein name

Locus Name

Acc#

gp:AB008550

AB008550

Description

Pseudomonas aeruginosa phage phi CTX, complete genome sequence.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
6834807...f1...78.....	487	5709	130	393	127	3.1e-08

Protein name

Locus Name

Acc#

probable dnaK suppressor

pir:D71366

D71366

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
817827...c3...515.....	488	5710	253	762	283	9.0e-25

Protein name

Locus Name

Acc#

rRNA methylase homolog ysgA

pir:G69984

G69984

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
829436_c3_535	489	5711	343	1032	600	2.3e-58

Protein name

Locus Name

Acc#

protein kinase homolog Thi

gp:AF070520

AF070520

Description

Sinorhizobium meliloti protein kinase homolog Thi (thi) andExoP-like protein genes, complete cds; and unknown genes.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
84637_c1_366	490	5712	68	207		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
9884427_f3_287	491	5713	66	201		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
9957827_c1_326	492	5714	413	1242	110	3.3e-14

Protein name

Locus Name

Acc#

sp:YBGH_ECOLI

P75742

Description

HYPOTHETICAL 54.2 KD PROTEIN IN PHRB-NEI INTERGENIC REGION

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
10320312_f2_49	493	5715	328	987		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
10681577_c3_251	494	5716	103	312	114	7.3e-07

Protein name

Locus Name

Acc#

hypothetical protein APE1165

pir:H72586

H72586

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
10828567_f3_89	495	5717	204	615	148	1.8e-10

Protein name

Locus Name

Acc#

conserved hypothetical protein

pir:C72361

C72361

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
13865887_c2_187	496	5718	64	195		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
1414187_f3_80	497	5719	316	951	357	1.3e-32

Protein name

Locus Name

Acc#

gp:AB012956

AB012956

Description

Vibrio cholerae genes for O-antigen synthesis, strain MO45, complete cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
14485841_f3_95	498	5720	208	627	659	1.3e-64
Protein name			Locus Name		Acc#	
rubrerythrin			gp:AF202316		AF202316	
Description						
Moorella thermoacetica rubrerythrin gene, complete cds.						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
14491537_f1_22	499	5721	682	2049	101	0.012
Protein name			Locus Name		Acc#	
comeA protein-related protein			pir:F72301		F72301	
Description						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
14569387_f3_81	500	5722	158	477		
Protein name			Locus Name		Acc#	
Description						
NO-HIT						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
15105001_c2_156	501	5723	119	360	87	0.013
Protein name			Locus Name		Acc#	
hypothetical protein M70.1			pir:T33032		T33032	
Description						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
15751503_c3_228	502	5724	219	660		
Protein name			Locus Name		Acc#	
Description						
NO-HIT						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
158136_f3_76	503	5725	308	927	399	4.6e-37

Protein name	Locus Name	Acc#
conserved hypothetical protein	pir:G72409	G72409

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
16056463_c3_236	504	5726	75	228		

Protein name	Locus Name	Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
194025_f1_30	505	5727	145	438		

Protein name	Locus Name	Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
20506875_f3_79	506	5728	211	636	202	3.5e-16

Protein name	Locus Name	Acc#
	sp:Y516_BORBU	O51468

Description

HYPOTHETICAL TRNA/RRNA METHYLTRANSFERASE BB0516,

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
20509632_f1_31	507	5729	448	1347	829	1.2e-82

Protein name	Locus Name	Acc#
dihydrolipoamide dehydrogenase, :2-oxoglutarate dehydrogenase complex chain E3:acetoin dehydrogenase complex	pir:I40794	I40794

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
20995143_c2_182	508	5730	83	252		

Protein name	Locus Name	Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
2220010_c1_147	509	5731	82	249		

Protein name	Locus Name	Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
22382311_f1_3	510	5732	318	957	287	3.4e-25

Protein name	Locus Name	Acc#
putative oxidoreductase	gp:SCF76	AL121600

Description

Streptomyces coelicolor cosmid F76.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

22679637_f1_11	511	5733	199	600	366	1.4e-33
----------------	-----	------	-----	-----	-----	---------

Protein name

Locus Name

Acc#

conserved hypothetical protein ysnA

pir:C69986

C69986

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

23516942_f1_15	512	5734	493	1482	127	8.3e-05
----------------	-----	------	-----	------	-----	---------

Protein name

Locus Name

Acc#

outer membrane protein tolC precursor (tolC)
RP224

pir:H71733

H71733

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

23594000_f1_17	513	5735	192	579		
----------------	-----	------	-----	-----	--	--

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

23632911_c2_157	514	5736	506	1521	1269	3.0e-129
-----------------	-----	------	-----	------	------	----------

Protein name

Locus Name

Acc#

sp:YGFH_ECOLI

P52043

Description

HYPOTHETICAL 53.8 KD PROTEIN IN SBM-FBA INTERGENIC REGION (O492)

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
23986267_f1_19	515	5737	148	447		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24116567_f2_53	516	5738	457	1374	194	2.5e-12

Protein name

Locus Name

Acc#

chromosomal hemolysin D

gp:AF081284

AF081284

Description

Escherichia coli strain CFT073 chromosomal hemolysin D (hlyD) gene, partial cds; and Hp1 (hp1), Hp2 (hp2), Hp3 (hp3), and Hp4 (hp4) genes, complete cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24225302_f1_16	517	5739	607	1824		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24260952_c3_221	518	5740	84	255		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24409662_c2_166	519	5741	104	315	107	6.8e-11
<u>Protein name</u>			<u>Locus Name</u>			<u>Acc#</u>
iron-uptake factor			gp:AF051690			AF051690
<u>Description</u>						
Pseudomonas aeruginosa iron-uptake factor (piuC), hydroxamate-typeferrisiderophore receptor (piuA), and iron-uptake factor (piuB) genes, complete cds.						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24415875_f2_55	520	5742	538	1617	521	7.2e-68
<u>Protein name</u>			<u>Locus Name</u>			<u>Acc#</u>
arylsulfatase			gp:PAATSAGN			Z48540
<u>Description</u>						
Pseudomonas aeruginosa atsR, atsB, atsC & atsA genes.						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24495337_c2_190	521	5743	338	1017	962	1.0e-96
<u>Protein name</u>			<u>Locus Name</u>			<u>Acc#</u>
			sp:NADA_SYNY3			P74578
<u>Description</u>						
QUINOLINATE SYNTHETASE A						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24641537_f3_104	522	5744	474	1425	417	5.7e-39
<u>Protein name</u>			<u>Locus Name</u>			<u>Acc#</u>
			sp:FUCO_RAT			P17164
<u>Description</u>						
I) (ALPHA-L-FUCOSIDE FUCOHYDROLASE)						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

24713961_f2_37	523	5745	304	915	376	1.3e-34
----------------	-----	------	-----	-----	-----	---------

Protein name

Locus Name

Acc#

prolipoprotein diacylglycerol transferase
(lgt) RP046

pir:F71712

F71712

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

25505386_f1_4	524	5746	244	735	232	2.3e-19
---------------	-----	------	-----	-----	-----	---------

Protein name

Locus Name

Acc#

chloramphenicol acetyltransferase

gp:AF124757

AF124757

Description

Zymomonas mobilis fosmid clone 43D2, complete sequence.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

25579383_f3_82	525	5747	627	1884	81	0.0020
----------------	-----	------	-----	------	----	--------

Protein name

Locus Name

Acc#

sp:EREB_ECOLI

P05789

Description

ERYTHROMYCIN ESTERASE TYPE II,

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

26367135_f3_70	526	5748	374	1125	1148	2.0e-116
----------------	-----	------	-----	------	------	----------

Protein name

Locus Name

Acc#

sp:YYAF_BACSU

P37518

Description

REGION

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

26757637_f3_88	527	5749	735	2208	867	1.2e-86
----------------	-----	------	-----	------	-----	---------

Protein name

Locus Name

Acc#

hemolysin secretion protein hlyB:protein
sll1180:protein sll1180

pir:S75806

S75806

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

2741426_f3_98	528	5750	397	1194	355	2.1e-32
---------------	-----	------	-----	------	-----	---------

Protein name

Locus Name

Acc#

sp:PBP_BACSU

P39844

Description

PUTATIVE PENICILLIN BINDING PROTEIN PRECURSOR

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

2928307_c3_215	529	5751	527	1584	1185	2.4e-120
----------------	-----	------	-----	------	------	----------

Protein name

Locus Name

Acc#

sp:NADB_PSEAE

Description

L-ASPARTATE OXIDASE, (QUINOLINATE SYNTHETASE B)

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

33992307_f1_23	530	5752	261	786	257	5.1e-22
----------------	-----	------	-----	-----	-----	---------

Protein name

Locus Name

Acc#

sp:Y117_HELPY

P56080

Description

HYPOTHETICAL PROTEIN HP0117

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
34376678_c1_141	531	5753	892	2679	1603	1.2e-164
Protein name			Locus Name			Acc#
			sp:MUTS_HAEIN			P44834
Description						
DNA MISMATCH REPAIR PROTEIN MUTS						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4094128_f2_51	532	5754	599	1800	83	0.026
Protein name			Locus Name			Acc#
erythromycin esterase homolog ybfO			pir:A69750			A69750
Description						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
424042_f3_103	533	5755	470	1413	557	1.2e-101
Protein name			Locus Name			Acc#
putative protein			gp:ATAP22			Z99708
Description						
Arabidopsis thaliana DNA chromosome 4, ESSA I AP2 contig fragmentNo. 2.						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4332837_f3_86	534	5756	228	687		
Protein name			Locus Name			Acc#
Description						
NO-HIT						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

4422762_f2_52	535	5757	238	717	172	8.4e-12
---------------	-----	------	-----	-----	-----	---------

Protein name

Locus Name

Acc#

putative glucosyl transferase

gp:AF105116

AF105116

Description

Streptococcus pneumoniae type 19C Cps19CR (cps19CR) gene, partialcds; putative oligosaccharide repeat unit transporter (cps19CJ), UDP-N-acetyl glucosamine-2-epimerase (cps19CK), and putativeglucosyl transferase (cps19CS) genes, complete cds; andglucose-1-phosphate thymidyl transferase (cps19CL) gene, partialcds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

4547063_c2_181	536	5758	94	285		
----------------	-----	------	----	-----	--	--

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

4569126_f2_68	537	5759	88	267	87	0.0057
---------------	-----	------	----	-----	----	--------

Protein name

Locus Name

Acc#

sp:PBP4_HAEIN

P45161

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

4901587_c1_114	538	5760	562	1689	1101	1.9e-111
----------------	-----	------	-----	------	------	----------

Protein name

Locus Name

Acc#

probable sulfate transporter

pir:A71463

A71463

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
5985875_c3_220	539	5761	810	2433	615	5.9e-60
<u>Protein name</u>			<u>Locus Name</u>		<u>Acc#</u>	
ferrichrome-iron receptor 3:protein slr1490:protein slr1490			pir:S74457		S74457	
<u>Description</u>						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
6.76.7057...f2_34.....	540	5762	372	1119	166	3.7e-09
<u>Protein name</u>			<u>Locus Name</u>		<u>Acc#</u>	
hypothetical protein PAB1767			pir:B75136		B75136	
<u>Description</u>						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
8.13.302...f1_9.....	541	5763	373	1122		
<u>Protein name</u>			<u>Locus Name</u>		<u>Acc#</u>	
<u>Description</u>			NO-HIT			

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
9.70680...f1_10.....	542	5764	965	2898	1588	5.1e-217
<u>Protein name</u>			<u>Locus Name</u>		<u>Acc#</u>	
putative leucyl tRNA synthetase			gp:AF069441		AF069441	
<u>Description</u>						
Arabidopsis thaliana BAC T19B17 from chromosome IV, near 19.3 cM, complete sequence.						

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
10158452_c2_398	543	5765	250	753	587	5.5e-57

Protein name

Locus Name

Acc#

putative glycosyl transferase

gp:AF048749

AF048749

Description

Bacteroides fragilis capsular polysaccharide biosynthesis operon, complete sequence.

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
1054637_f3_214	544	5766	211	636	1033	3.0e-104

Protein name

Locus Name

Acc#

superoxide dismutase

gp:BNRSOD2

D13756

Description

Bacteroides fragilis DNA for superoxide dismutase, complete cds.

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
10750067_c2_445.....	545	5767	126	381		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
10803580_c2_394.....	546	5768	283	852	803	7.1e-80

Protein name

Locus Name

Acc#

alpha-D-glucose-1-phosphate

gp:YEPASCA

L27130

Description

Yersinia pseudotuberculosis alpha-D-glucose-1-phosphatecytidyltransferase (asca) gene, complete cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

10837887_c1_303	547	5769	374	1125	1002	5.8e-101
-----------------	-----	------	-----	------	------	----------

Protein name

Locus Name

Acc#

CDP-glucose-4,6-dehydratase

pir:D47070

D47070

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

10978425_c3_473	548	5770	61	186		
-----------------	-----	------	----	-----	--	--

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

11214032_c2_377	549	5771	477	1434	837	1.8e-83
-----------------	-----	------	-----	------	-----	---------

Protein name

Locus Name

Acc#

sp:ATOC_ECOLI

Q06065

Description

DECARBOXYLASE INHIBITOR) (ORNITHINE DECARBOXYLASE ANTIZYME)

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

11932290_f2_88	550	5772	107	324	152	1.2e-10
----------------	-----	------	-----	-----	-----	---------

Protein name

Locus Name

Acc#

sp:CBIK_SALTY

Q05592

Description

CBIK PROTEIN

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
126376_f3_197	551	5773	500	1503		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
13.785926_c2_3.72.....	552	5774	172	519		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
13834812_c3_485.....	553	5775	347	1044		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
14461567_f3_222.....	554	5776	386	1161	890	4.3e-89

Protein name

Locus Name

Acc#

ThiH

gp:AF154064

AF154064

Description

Salmonella typhimurium ThiH (thiH) gene, complete cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

14489050_f2_180	555	5777	555	1668	147	4.7e-07
-----------------	-----	------	-----	------	-----	---------

Protein name

Locus Name

Acc#

aspartate aminotransferase

pir:D75496

D75496

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

14642207_f1_18	556	5778	595	1788	1878	8.6e-194
----------------	-----	------	-----	------	------	----------

Protein name

Locus Name

Acc#

sp:THIC_BACSU

Description

THIAMINE BIOSYNTHESIS PROTEIN THIC

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

14647206_f3_237	557	5779	257	774	96	0.021
-----------------	-----	------	-----	-----	----	-------

Protein name

Locus Name

Acc#

conserved hypothetical protein MTH469

pir:D69161

D69161

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

14729186_f2_148	558	5780	155	468	88	0.029
-----------------	-----	------	-----	-----	----	-------

Protein name

Locus Name

Acc#

sp:GENK_ECOLI

P02988

Description

PROTEIN K

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
14867327_f3_236	559	5781	408	1227	175	3.0e-10

Protein name

Locus Name

Acc#

sp:YIGN_ECOLI

P27850

Description

HYPOTHETICAL 54.7 KD PROTEIN IN UDP-UBIE INTERGENIC REGION PRECURSOR

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
16064015_c2_385	560	5782	140	423	84	0.0060

Protein name

Locus Name

Acc#

trbA protein

pir:A49852

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
164043_c1_309.....	561	5783	251	756	493	5.0e-47

Protein name

Locus Name

Acc#

conserved hypothetical protein HP0162

pir:B64540

B64540

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
194086_c3_520.....	562	5784	790	2373	1337	2.8e-144

Protein name

Locus Name

Acc#

sp:PCRA_BACST

P56255

Description

ATP-DEPENDENT HELICASE PCRA,

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

1960876_c2_403	563	5785	72	219	81	0.0023
----------------	-----	------	----	-----	----	--------

Protein name

Locus Name

Acc#

hypothetical protein MJ1608

pir:G64500

G64500

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

19689678_c1_307	564	5786	442	1329	128	5.0e-05
-----------------	-----	------	-----	------	-----	---------

Protein name

Locus Name

Acc#

unknown

gp:AF144879

AF144879

Description

Leptospira interrogans rfb locus, complete sequence.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

19709682_c1_302	565	5787	451	1356	1278	3.3e-130
-----------------	-----	------	-----	------	------	----------

Protein name

Locus Name

Acc#

CDP-4-keto-6-deoxy-D-glucose-3-dehydratase

gp:YPE251713

AJ251713

Description

Yersinia pestis strain EV76 hemH gene (partial) and O-antigen genecluster for ddhD gene, ddhA gene, ddhB pseudogene, ddhC gene, prtgene, wbyH gene, wzx gene, wbyI pseudogene, wbyJ gene, wzypseudogene, wbyK gene, gmd pseudogene, fcl pseudogene, manC gene, wbyL gene, manB gene, wzz gene and gsk gene (partial).

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

19711067_c3_486	566	5788	300	903	374	2.1e-34
-----------------	-----	------	-----	-----	-----	---------

Protein name

Locus Name

Acc#

hypothetical protein jhp0094

pir:E71975

E71975

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
20087751_c3_489	567	5789	318	957	1429	3.3e-146

Protein name

Locus Name

Acc#

putative UDP-GlcNAc:undecaprenylphosphate

gp:AF048749

AF048749

Description

Bacteroides fragilis capsular polysaccharide biosynthesis operon, complete sequence.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
20520302_c3_462	568	5790	495	1488	112	9.6e-06

Protein name

Locus Name

Acc#

immunoreactive 50kD antigen PG53

gp:AF175720

AF175720

Description

Porphyromonas gingivalis strain W50 immunoreactive 50kD antigenPG53 gene, complete cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
20596012..f1..1.....	569	5791	801	2406	181	9.9e-33

Protein name

Locus Name

Acc#

ferrichrome-iron receptor 3:protein
slr1490:protein slr1490

pir:S74457

S74457

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
21495928..f1..23.....	570	5792	94	285		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
-----------------	-------------	-------------	------------------	------------------	--------------	--------------------

2150305_c1_308	571	5793	299	900	174	1.1e-20
----------------	-----	------	-----	-----	-----	---------

Protein name

Locus Name

Acc#

UDP-glucose-4-epimerase/dTDP-glucose-4,6

gp:AF048749

AF048749

Description

Bacteroides fragilis capsular polysaccharide biosynthesis operon, complete sequence.

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
-----------------	-------------	-------------	------------------	------------------	--------------	--------------------

22114755_f1_7	572	5794	478	1437	384	1.8e-35
---------------	-----	------	-----	------	-----	---------

Protein name

Locus Name

Acc#

precorrin-6Y methylase:protein
sll0099:protein sll0099

pir:S76697

S76697

Description

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
-----------------	-------------	-------------	------------------	------------------	--------------	--------------------

22129152_f2_112.....	573	5795	632	1899		
----------------------	-----	------	-----	------	--	--

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
-----------------	-------------	-------------	------------------	------------------	--------------	--------------------

2349150_c1_344.....	574	5796	103	312		
---------------------	-----	------	-----	-----	--	--

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
2351432_f2_175	575	5797	68	207		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
23562762_c3_488	576	5798	240	723	366	1.4e-33

Protein name

Locus Name

Acc#

putative glycosyl transferase

gp:AF071085

AF071085

Description

Enterococcus faecalis strain OG1RF polysaccharide biosynthetic genecluster, partial sequence.

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
23611383_f1_38	577	5799	68	207		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
23615905_c3_523	578	5800	409	1230	327	6.3e-32

Protein name

Locus Name

Acc#

sp:HXX3_HUMAN

P52790

Description

HEXOKINASE TYPE III, (HK III)

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
23617802_f2_160	579	5801	82	249	84	0.0054
Protein name			Locus Name		Acc#	
PP31 39K orf36			pir:T41782		T41782	
Description						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
23704688_c3_459.....	580	5802	439	1320	132	1.8e-11
Protein name			Locus Name		Acc#	
conserved hypothetical protein yknZ			pir:E69858		E69858	
Description						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
23831325_f2_155.....	581	5803	95	288		
Protein name			Locus Name		Acc#	
Description			NO-HIT			

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
23851567_f2_119.....	582	5804	192	579		
Protein name			Locus Name		Acc#	
Description			NO-HIT			

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
23851637_f3_217.....	583	5805	266	801	824	4.2e-82
Protein name			Locus Name		Acc#	
Description			sp:THIG_ECOLI			
THIG PROTEIN						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
23957812_f3_267	584	5806	109	330	91	0.0011

Protein name

Locus Name

Acc#

chaperone GrpE type 2

gp:AF098636

AF098636

Description

Nicotiana tabacum chaperone GrpE type 2 (GrpE2) mRNA, nuclear gene encoding mitochondrial protein, complete cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24023462_c1_311	585	5807	426	1281	618	2.9e-60

Protein name

Locus Name

Acc#

sp:YDAR_BACSU

P96593

Description

HYPOTHETICAL 45.7 KD PROTEIN IN MUTT-GSIB INTERGENIC REGION

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24035952_f2_147.....	586	5808	95	288	82	0.0018

Protein name

Locus Name

Acc#

unknown protein

gp:SCCXV106K

X95258

Description

S.cerevisiae 10.6kbp fragment from chromosome XV.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24223762_c1_318.....	587	5809	448	1347	503	1.6e-82

Protein name

Locus Name

Acc#

Na⁺/H⁺-exchanging protein:Na⁺/H⁺ antiporter

pir:JX0360

JX0360

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

24239006_f3_269	588	5810	337	1014	110	0.0058
-----------------	-----	------	-----	------	-----	--------

Protein name

Locus Name

Acc#

gp:ECORHSEX

L19083

Description

Escherichia coli RhSE genetic element; defective RhSE core protein, complete cds; complete ORF-E2; H-rpt subelement; complete ORF-H.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

24303127_f2_173	589	5811	142	429		
-----------------	-----	------	-----	-----	--	--

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

24407687_c2_400	590	5812	827	2484	1188	1.1e-120
-----------------	-----	------	-----	------	------	----------

Protein name

Locus Name

Acc#

sp:SYFB_ECOLI

Description

TRNA LIGASE BETA CHAIN) (PHERS)

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

24410780_f1_71	591	5813	63	192		
----------------	-----	------	----	-----	--	--

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
24412912_f2_94	592	5814	192	579		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
245047113_f3_200.....	593	5815	240	723	291	1.3e-25

Protein name

Locus Name

Acc#

hypothetical protein MTH671

pir:D69189

D69189

Description

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
24614415_f1_33.....	594	5816	315	948	337	1.7e-30

Protein name

Locus Name

Acc#

sp:YLYB_BACSU

Description

HYPOTHETICAL 33.7 KD PROTEIN IN LSP-PYRR INTERGENIC REGION (ORF-X)

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
24641937_f1_8.....	595	5817	601	1806	628	2.5e-61

Protein name

Locus Name

Acc#

precorrin-3 methylase

pir:A64497

A64497

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24642311_c2_392	596	5818	162	489	223	2.1e-18
Protein name			Locus Name			Acc#
unknown			gp:AF048749			AF048749
Description						

Bacteroides fragilis capsular polysaccharide biosynthesis operon, complete sequence.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24648563_c1_322	597	5819	136	411		
Protein name			Locus Name			Acc#
Description						

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24694013_f1_3.....	598	5820	1326	3981	659	1.8e-115
Protein name			Locus Name			Acc#
cobalamin biosynthesis protein N			pir:C69048			C69048
Description						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
25906675_c3_492.....	599	5821	173	522	106	3.1e-05
Protein name			Locus Name			Acc#
hypothetical protein AF0456			pir:H69306			H69306
Description						

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
25976708_f3_209	600	5822	95	288		

<u>Protein name</u>	<u>Locus Name</u>	<u>Acc#</u>

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
26362791_f3_272	601	5823	603	1812	683	3.7e-67

<u>Protein name</u>	<u>Locus Name</u>	<u>Acc#</u>

probable membrane protein b0847

pir:G64822

G64822

Description

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
26604510_c1_314	602	5824	595	1788	1902	2.5e-196

<u>Protein name</u>	<u>Locus Name</u>	<u>Acc#</u>

sp:LEPA_BACSU

P37949

Description

GTP-BINDING PROTEIN LEPA

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
26835887_f1_20	603	5825	238	717	518	1.1e-49

<u>Protein name</u>	<u>Locus Name</u>	<u>Acc#</u>

MPT-synthase sulfurylase

gp:SYPCCMOEB

Y16560

Description

Synechococcus PCC7942 moeB gene.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
2853436_c3_500	604	5826	148	447		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
29353952_c1_304	605	5827	592	1779	891	3.4e-89

Protein name

Locus Name

Acc#

gp:AF025396

AF025396

Description

Vibrio anguillarum rfb region, partial sequence.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
30522966_f2_122	606	5828	262	789	132	5.9e-07

Protein name

Locus Name

Acc#

sp:TEN1_BACSU

P25053

Description

REGULATORY PROTEIN TEN1

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
32165905_f1_5	607	5829	109	330	82	0.0018

Protein name

Locus Name

Acc#

hypothetical protein MTH670

pir:C69189

C69189

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
33289500_c2_393	608	5830	307	924	1394	1.7e-142
<u>Protein name</u>			<u>Locus Name</u>		<u>Acc#</u>	
glucose-1-phosphate thymidyl transferase			gp:AF048749		AF048749	
<u>Description</u>						
Bacteroides fragilis capsular polysaccharide biosynthesis operon, complete sequence.						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
34064010_f3_232	609	5831	478	1437	707	1.1e-69
<u>Protein name</u>			<u>Locus Name</u>		<u>Acc#</u>	
RNA methyltransferase homolog yefA			pir:E69793		E69793	
<u>Description</u>						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
34079635_f1_9.....	610	5832	168	507		
<u>Protein name</u>			<u>Locus Name</u>		<u>Acc#</u>	
<u>Description</u>						
NO-HIT						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
34173556_c3_482.....	611	5833	314	945	223	4.7e-18
<u>Protein name</u>			<u>Locus Name</u>		<u>Acc#</u>	
ADP-L-glycero-D-manno-heptose-6-epimerase			pir:G70330		G70330	
<u>Description</u>						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
34245941_c3_490	612	5834	86	261		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
34384631_f3_216.....	613	5835	205	618	319	1.4e-28

Protein name

Locus Name

Acc#

sp:THIE_SYNY3

P72965

Description

PYROPHOSPHORYLASE) (TMP-PPASE) (THIAMINE-PHOSPHATE SYNTHASE)

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
35214385_c2_397.....	614	5836	297	894	182	6.8e-13

Protein name

Locus Name

Acc#

glucosyl transferase

gp:SMU52844

U52844

Description

Serratia marcescens putative glycosyltransferase, putativeglycosyltransferase, putative heptosylIII transferase (waaQ), 3-deoxy-manno-octulosonic acid transferase (waaA), glucosyltransferase (waaE), and KdtB (kdtB) genes, complete cds; and Fpg(fpg) gene, partial cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
35348458_c1_327.....	615	5837	119	360	93	0.00025

Protein name

Locus Name

Acc#

unknown

gp:AF007381

AF007381

Description

Flavobacterium johnsoniae gliding motility protein (gldA) gene, complete cds; and unknown genes.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
36330078_c3_518	616	5838	96	291		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
36601442...f3_215.....	617	5839	118	357	118	2.8e-07

Protein name

Locus Name

Acc#

gp:STYSTMF1

AF170176

Description

Salmonella typhimurium fragment STMF1.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
3939512_c1_326.....	618	5840	909	2730	1830	4.0e-281

Protein name

Locus Name

Acc#

sp:PODK_CLOSY

P22983

Description

DIKINASE)

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
3940877_f2_96.....	619	5841	472	1419	591	9.3e-60

Protein name

Locus Name

Acc#

precorrin-3 methylase

gp:BMAJ758

AJ000758

Description

Bacillus megaterium 16kb genomic sequence, cobalamin operon.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
39642_c3_480	620	5842	196	591	841	6.7e-84
Protein name			Locus Name		Acc#	
dTDP-6-deoxy-D-glucose-3,5 epimerase			gp:AF048749		AF048749	
Description						
Bacteroides fragilis capsular polysaccharide biosynthesis operon, complete sequence.						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
40930_f2_144	621	5843	140	423	391	3.2e-36
Protein name			Locus Name		Acc#	
conserved hypothetical protein			pir:C75256		C75256	
Description						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4196081_c2_391.....	622	5844	156	471		
Protein name			Locus Name		Acc#	
Description						
NO-HIT						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4334455_c3_463.....	623	5845	436	1311	227	3.5e-19
Protein name			Locus Name		Acc#	
			sp:NTRB_RHOCA		P09431	
Description						
NITROGEN REGULATION PROTEIN NTRB,						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4688828_c3_507	624	5846	96	291	87	0.0012

Protein name

Locus Name

Acc#

unknown

gp:AF007381

AF007381

Description

Flavobacterium johnsoniae gliding motility protein (gldA) gene, complete cds; and unknown genes.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4881512_f1_72	625	5847	153	462		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4881557_c2_404	626	5848	395	1188	105	0.016

Protein name

Locus Name

Acc#

Na⁺/H⁺-exchanging protein sll0689:Na⁺/H⁺ antiporter:Na⁺/H⁺ antiporter

pir:S74414

S74414

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4882755_f3_268	627	5849	311	936	103	0.0030

Protein name

Locus Name

Acc#

growth-associated protein

gp:ZEF GAP

L27645

Description

Brachydanio rerio growth-associated protein, complete cds.

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
4884635_c3_487	628	5850	266	801	403	1.7e-37

<u>Protein name</u>	<u>Locus Name</u>	<u>Acc#</u>
unknown	gp:AF144879	AF144879

Description

Leptospira interrogans rfb locus, complete sequence.

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
4884712_c2_401	629	5851	254	765	636	3.5e-62

<u>Protein name</u>	<u>Locus Name</u>	<u>Acc#</u>
exodeoxyribonuclease	pir:B69126	B69126

Description

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
4957512_c3_506.....	630	5852	193	582		

<u>Protein name</u>	<u>Locus Name</u>	<u>Acc#</u>

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
504757_c2_373.....	631	5853	956	2871	383	3.0e-34

<u>Protein name</u>	<u>Locus Name</u>	<u>Acc#</u>
RcsC	gp:AF071215	AF071215

Description

Proteus mirabilis regulator of swarming behavior precursor (rsbA) and RcsB (rbsB) genes, complete cds; and RcsC (rscC) gene, partialcds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

5110325_c3_521	632	5854	381	1146	1054	1.8e-106
----------------	-----	------	-----	------	------	----------

Protein name

Locus Name

Acc#

carboxynorspermidine decarboxylase:protein
sll0873:protein sll0873

pir:S77268

S77268

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

5112802_f3_207	633	5855	644	1935	423	2.8e-39
----------------	-----	------	-----	------	-----	---------

Protein name

Locus Name

Acc#

CbiD protein

gp:BMAJ758

AJ000758

Description

Bacillus megaterium 16kb genomic sequence, cobalamin operon.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

6025675_f2_162	634	5856	75	228		
----------------	-----	------	----	-----	--	--

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

6913875_c1_305	635	5857	450	1353	602	1.4e-58
----------------	-----	------	-----	------	-----	---------

Protein name

Locus Name

Acc#

hypothetical protein

pir:S22614

S22614

Description

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
7087642_c1_306	636	5858	61	186		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
910262_f3_235.....	637	5859	305	918	670	8.8e-66

Protein name

Locus Name

Acc#

sp:AMP1_SYNY3

P53579

Description

M)

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
969165_c1_297.....	638	5860	336	1011	182	1.1e-11

Protein name

Locus Name

Acc#

conserved hypothetical protein MTH1261

pir:F69035

F69035

Description

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
9885166_c3_460.....	639	5861	282	849	105	0.014

Protein name

Locus Name

Acc#

sp:YBJZ_ECOLI

P75831

Description

HYPOTHETICAL ABC TRANSPORTER ATP-BINDING PROTEIN YBJZ

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
11757880_c2_81	640	5862	146	441	108	3.2e-06
Protein name			Locus Name		Acc#	
hypothetical protein PH1670			pir:F71047		F71047	
Description						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
14460957_f1_4	641	5863	363	1092		
Protein name			Locus Name		Acc#	
Description						
NO-HIT						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
14728413_f1_8	642	5864	95	288		
Protein name			Locus Name		Acc#	
Description						
NO-HIT						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
15878777_c3_95	643	5865	232	699	178	1.1e-12
Protein name			Locus Name		Acc#	
serine-rich protein			pir:T39903		T39903	
Description						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
20741053_f2_25	644	5866	240	723	1219	5.9e-124
Protein name			Locus Name		Acc#	
BatC			gp:AF116251		AF116251	
Description						

Bacteroides fragilis batI operon, complete sequence.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
209688_f3_38	645	5867	63	192		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
22027_f3_31	646	5868	355	1068	683	3.7e-67

Protein name

Locus Name

Acc#

sp:GCP_HAEIN

P43764

Description

(GLYCOPROTEASE)

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
22039692_f2_24	647	5869	216	651	1014	3.1e-102

Protein name

Locus Name

Acc#

BatB

gp:AF116251

AF116251

Description

Bacteroides fragilis batI operon, complete sequence.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
22297140_f1_3	648	5870	327	984	768	3.6e-76

Protein name

Locus Name

Acc#

sp:FTSY_HAEIN

P44870

Description

CELL DIVISION PROTEIN FTSY

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
23649052_f1_9	649	5871	399	1200	166	1.2e-11

Protein name

Locus Name

Acc#

sp:Y531_METJA

Q57951

Description

HYPOTHETICAL PROTEIN MJ0531

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
23834376_c2_84	650	5872	122	369	122	1.0e-07

Protein name

Locus Name

Acc#

hypothetical protein APE1982

pir:H72500

H72500

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24259700_f1_6	651	5873	615	1848	3076	0.0

Protein name

Locus Name

Acc#

BatD

gp:AF116251

AF116251

Description

Bacteroides fragilis batI operon, complete sequence.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24407537_f2_21	652	5874	93	282	177	1.5e-13

Protein name

Locus Name

Acc#

ribosomal protein L28

pir:E64104

E64104

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24415903_f1_7	653	5875	279	840	1381	4.0e-141
<u>Protein name</u>			<u>Locus Name</u>		<u>Acc#</u>	
BatE			gp:AF116251		AF116251	
<u>Description</u>						
Bacteroides fragilis batI operon, complete sequence.						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24823562_f1_2	654	5876	68	207	157	2.0e-11
<u>Protein name</u>			<u>Locus Name</u>		<u>Acc#</u>	
			sp:RK33_ODOS1		P49565	
<u>Description</u>						
CHLOROPLAST 50S RIBOSOMAL PROTEIN L33						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
25975010_c1_63.....	655	5877	1129	3390		
<u>Protein name</u>			<u>Locus Name</u>		<u>Acc#</u>	
<u>Description</u>						
NO-HIT						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
272212_f3_33.....	656	5878	492	1479	201	1.6e-12
<u>Protein name</u>			<u>Locus Name</u>		<u>Acc#</u>	
antigen 332			pir:JN0292		JN0292	
<u>Description</u>						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
3149069_c1_40	657	5879	847	2544	4304	0.0

Protein name	Locus Name	Acc#
DNA gyrase A subunit	gp:AB017712	AB017712

Description

Bacteroides fragilis gyrA gene for DNA gyrase A subunit, completecds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
33397127_f1_5	658	5880	331	996	1656	2.9e-170

Protein name	Locus Name	Acc#
BatA	gp:AF116251	AF116251

Description

Bacteroides fragilis batI operon, complete sequence.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
34101702_f2_23.....	659	5881	289	870	388	6.7e-36

Protein name	Locus Name	Acc#
conserved hypothetical protein BB0175	pir:G70121	G70121

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
34266968_f3_32.....	660	5882	454	1365	942	1.3e-94

Protein name	Locus Name	Acc#
conserved hypothetical protein aq_849	pir:E70373	E70373

Description

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
34564376_f1_1	661	5883	415	1248	603	1.1e-58

<u>Protein name</u>	<u>Locus Name</u>	<u>Acc#</u>
hypothetical protein	pir:S76561	S76561
<u>Description</u>		

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
3912925_f3_34	662	5884	334	1005	821	8.8e-82

<u>Protein name</u>	<u>Locus Name</u>	<u>Acc#</u>
probable moxR protein	pir:B70874	B70874
<u>Description</u>		

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
4961537_c3_87	663	5885	418	1257	164	6.9e-09

<u>Protein name</u>	<u>Locus Name</u>	<u>Acc#</u>
conserved hypothetical protein aq_854	pir:B70374	B70374
<u>Description</u>		

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
5345637_f2_22	664	5886	100	303	154	4.2e-11

<u>Protein name</u>	<u>Locus Name</u>	<u>Acc#</u>
	sp:DBH_THEMEA	P36206
<u>Description</u>		

DNA-BINDING PROTEIN HU

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
7072675_f3_36	665	5887	146	441	687	1.4e-67
Protein name			Locus Name		Acc#	
BatB			gp:AF116251		AF116251	
Description						
Bacteroides fragilis batI operon, complete sequence.						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
10562517_f3_78	666	5888	71	216		
Protein name			Locus Name		Acc#	
Description						
NO-HIT						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
12129682_f2_67	667	5889	516	1551	319	2.6e-26
Protein name			Locus Name		Acc#	
lipase-like protein			pir:A64706		A64706	
Description						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
12679062_f3_97	668	5890	66	201		
Protein name			Locus Name		Acc#	
Description						
NO-HIT						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
13834500_c3_196	669	5891	144	435	203	2.7e-16
Protein name			Locus Name		Acc#	
hypothetical protein BB0530			pir:A70166		A70166	
Description						

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
17086686_f3_95	670	5892	268	807		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
20585963_f1_19.....	671	5893	413	1242	226	1.8e-16

Protein name

Locus Name

Acc#

hypothetical protein jhp1380

pir:G71815

G71815

Description

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
2151552_c2_146.....	672	5894	250	753	104	2.7e-05

Protein name

Locus Name

Acc#

cytochrome b

gp:AF017516

AF017516

Description

Bombus pascuorum cytochrome b (cytb) gene, mitochondrial gene encoding mitochondrial protein, partial cds.

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
22473516_c3_186.....	673	5895	500	1503	1054	1.8e-106

Protein name

Locus Name

Acc#

sp:CBIP_SALTY

Q05597

Description

COBYRIC ACID SYNTHASE

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24269180_f2_71	674	5896	400	1203	263	2.9e-21
Protein name			Locus Name		Acc#	
hypothetical protein jhpl379			pir:F71815		F71815	
Description						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24317806_f2_53	675	5897	321	966	436	5.5e-41
Protein name			Locus Name		Acc#	
nicotinate-nucleotide--dimethylbenzimidazole phosphoribosyltransferase			pir:A75577		A75577	
Description						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24345167_f1_4	676	5898	206	621	310	1.2e-27
Protein name			Locus Name		Acc#	
cobinamide kinase / cobinamide phosphate guanylyltransferase			pir:S52220			
Description						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24417212_c3_197	677	5899	502	1509	1238	5.7e-126
Protein name			Locus Name		Acc#	
proline--tRNA ligase, proS:prolyl-tRNA synthetase:prolyl-tRNA synthetase			pir:A70150		A70150	
Description						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24641903_c2_166	678	5900	168	507		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24666005_c2_156.....	679	5901	114	345		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24822213_f2_68.....	680	5902	333	1002	664	3.8e-65

Protein name

Locus Name

Acc#

immunoreactive 36 kDa antigen PG14

gp:AF145798

AF145798

Description

Porphyromonas gingivalis strain W50 immunoreactive 36 kDa antigenPG14 gene, complete cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24822688_f2_66.....	681	5903	136	411	114	7.3e-07

Protein name

Locus Name

Acc#

hypothetical protein

pir:S76776

S76776

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
25401437_c1_115	682	5904	165	498	178	2.0e-13

Protein name

Locus Name

Acc#

sp:YJJP_HAEIN

P44520

Description

HYPOTHETICAL PROTEIN HI0108

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
30082887_c2_145	683	5905	279	840	238	5.3e-20

Protein name

Locus Name

Acc#

sp:YJJP_ECOLI

P39402

Description

HYPOTHETICAL 30.5 KD PROTEIN IN DNAT-BGLJ INTERGENIC REGION (F277)

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
31657080_c2_144	684	5906	580	1743	1106	5.5e-112

Protein name

Locus Name

Acc#

sp:YIDE_ECOLI

Description

HYPOTHETICAL 58.9 KD PROTEIN IN GLVC-IBPB INTERGENIC REGION (ORFA)

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
32228452_c3_172	685	5907	191	576	282	1.2e-24

Protein name

Locus Name

Acc#

conserved hypothetical protein yvqk

pir:D70046

D70046

Description

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
33642211_f1_9	686	5908	260	783	156	3.7e-11

<u>Protein name</u>	<u>Locus Name</u>	<u>Acc#</u>
probable phosphoglycerate mutase	pir:B75539	B75539

Description

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
34195888_c3_190	687	5909	325	978	489	1.3e-46

<u>Protein name</u>	<u>Locus Name</u>	<u>Acc#</u>
	sp:COBD_PSEDE	P21634

Description

COBD PROTEIN

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
422125_f1_16	688	5910	132	399	93	0.0029

<u>Protein name</u>	<u>Locus Name</u>	<u>Acc#</u>
beta-tropomyosin	pir:S23470	S23470

Description

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
4689027_c3_195	689	5911	1084	3255	895	1.3e-89

<u>Protein name</u>	<u>Locus Name</u>	<u>Acc#</u>
tricorn protease	gp:TAU72850	U72850

Description

Thermoplasma acidophilum GTP-binding protein and tricorn protease(TRI) genes, complete cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4807062_c1_116	690	5912	448	1347	657	2.1e-64

Protein name	Locus Name	Acc#
cobyrinic acid a,c-diamide synthase	pir:A75619	A75619

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
484451_c2_149	691	5913	821	2466	436	1.3e-37

Protein name	Locus Name	Acc#
two component sensor	gp:AF030352	AF030352

Description

Pseudomonas aeruginosa two component sensor (lemA) gene, partial cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
5116586_c1_131	692	5914	289	870	296	3.8e-26

Protein name	Locus Name	Acc#
CobD	gp:STU90625	U90625

Description

Salmonella typhimurium alpha-ribazole-5'-phosphate phosphatase CobC (cobC) gene, partial cds and putative aminotransferase CobD (cobD) gene, complete cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
5172508_f1_34	693	5915	103	312		

Protein name	Locus Name	Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
-----------------	-------------	-------------	------------------	------------------	--------------	--------------------

5181263_f2_54	694	5916	250	753	268	3.5e-23
---------------	-----	------	-----	-----	-----	---------

<u>Protein name</u>	<u>Locus Name</u>	<u>Acc#</u>
---------------------	-------------------	-------------

cobalamin synthase	pir:H75576	H75576
--------------------	------------	--------

Description

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
-----------------	-------------	-------------	------------------	------------------	--------------	--------------------

812510_c2_164	695	5917	66	201		
---------------	-----	------	----	-----	--	--

<u>Protein name</u>	<u>Locus Name</u>	<u>Acc#</u>
---------------------	-------------------	-------------

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
-----------------	-------------	-------------	------------------	------------------	--------------	--------------------

10042126_f3_172	696	5918	176	531	135	4.3e-09
-----------------	-----	------	-----	-----	-----	---------

<u>Protein name</u>	<u>Locus Name</u>	<u>Acc#</u>
---------------------	-------------------	-------------

hypothetical protein	gp:SSU18930	Y18930
----------------------	-------------	--------

Description

Sulfolobus solfataricus 281 kb genomic DNA fragment, strain P2.

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
-----------------	-------------	-------------	------------------	------------------	--------------	--------------------

11799076_c2_266	697	5919	103	312		
-----------------	-----	------	-----	-----	--	--

<u>Protein name</u>	<u>Locus Name</u>	<u>Acc#</u>
---------------------	-------------------	-------------

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
1199075_c2_263	698	5920	64	195		

Protein name Locus Name Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
1359507_c2_282	699	5921	434	1305	1073	1.7e-108

Protein name Locus Name Acc#

sp:HISX_ECOLI

Description

HISTIDINOL DEHYDROGENASE, (HDH)

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
1370937_c3_383	700	5922	119	360		

Protein name Locus Name Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
13756515_f2_119	701	5923	68	207		

Protein name Locus Name Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
13786251_f1_58	702	5924	636	1911	386	1.8e-33
Protein name			Locus Name		Acc#	
histidine kinase			gp:AF114442		AF114442	
Description						
Nostoc punctiforme histidine kinase (hepK) gene, complete cds.						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
14630063_c3_356	703	5925	389	1170	880	4.9e-88
Protein name			Locus Name		Acc#	
			sp:HIS7_HAEIN		P44327	
Description						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
15039087_c2_272	704	5926	938	2817	2717	1.1e-282
Protein name			Locus Name		Acc#	
B12-dependent			gp:ECOUW89		U00006	
Description						
E. coli chromosomal region from 89.2 to 92.8 minutes.						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
15085902_f2_125	705	5927	788	2367	1213	2.8e-129
Protein name			Locus Name		Acc#	
			sp:RHO_PSEFL		P52155	
Description						
TRANSCRIPTION TERMINATION FACTOR RHO						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
15629642_f2_79	706	5928	380	1143	113	0.0067

Protein name

Locus Name

Acc#

gp:PFMAL3P2

Description

Plasmodium falciparum MAL3P2, complete sequence.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
1995452_c3_340	707	5929	133	402	88	0.0023

Protein name

Locus Name

Acc#

gp:SYCPUR1

L36958

Description

Synechocystis sp. (clone pSYN411) glycineamide ribonucleotidetransformylase (purT), Orf134 and dnaA genes, complete cds, photosystem II reaction center protein D2 (psbD) gene, 5' end.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
20119062_f3_159	708	5930	642	1929	1069	1.1e-123

Protein name

Locus Name

Acc#

hypothetical protein Rv2438c

pir:D70680

D70680

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
21520006_c2_301	709	5931	289	870	101	0.0091

Protein name

Locus Name

Acc#

hypothetical protein

gp:AF021091

AF021091

Description

Helicobacter pylori hypothetical protein (HP0395), hypothetical protein (HP0394), chemotaxis protein CheV (cheV), bifunctional chemotaxis protein CheF (cheF), chemotaxis protein CheW (cheW), and adhesin-thiol peroxidase TagD (tagD) genes, complete cds; and superoxide dismutase SodB (sodB) gene, partial cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
2161286_f3_160	710	5932	192	579		

Protein name _____ Locus Name _____ Acc# _____

Description _____

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
21641562_c2_271	711	5933	198	597		

Protein name _____ Locus Name _____ Acc# _____

Description _____

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
22289781_f3_195	712	5934	676	2031	213	1.5e-13

Protein name _____ Locus Name _____ Acc# _____

sp:PLEC_CAUCR P37894

Description _____

NON-MOTILE AND PHAGE-RESISTANCE PROTEIN,

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
22902267_c2_318	713	5935	252	759		

Protein name _____ Locus Name _____ Acc# _____

Description _____

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
23477187_f1_54	714	5936	315	948	398	5.9e-37
Protein name			Locus Name		Acc#	
BrkB			pir:I40328		I40328	
Description						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
23522812_f2_70	715	5937	63	192		
Protein name			Locus Name		Acc#	
Description						

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
23593750_f2_84	716	5938	726	2181		
Protein name			Locus Name		Acc#	
Description						

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
2381338_c2_316	717	5939	347	1044	87	0.025
Protein name			Locus Name		Acc#	
hypothetical protein PH0161			pir:G71237		G71237	
Description						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24259427_c3_381	718	5940	265	798		
Protein name			Locus Name		Acc#	
Description						

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
24335943_c1_236	719	5941	385	1158	190	8.0e-12
<u>Protein name</u>			<u>Locus Name</u>		<u>Acc#</u>	
conserved hypothetical protein MTH884			pir:B69218		B69218	
<u>Description</u>						

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
24408517_c3_384	720	5942	93	282		
<u>Protein name</u>			<u>Locus Name</u>		<u>Acc#</u>	
<u>Description</u>						

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
24640677_c2_317	721	5943	290	873		
<u>Protein name</u>			<u>Locus Name</u>		<u>Acc#</u>	
<u>Description</u>						

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
24641080_c1_214	722	5944	136	411		
<u>Protein name</u>			<u>Locus Name</u>		<u>Acc#</u>	
<u>Description</u>						

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
24643887_c2_314	723	5945	355	1068		
<u>Protein name</u>			<u>Locus Name</u>		<u>Acc#</u>	
<u>Description</u>						

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24648538_c1_233	724	5946	187	564	331	7.4e-30

Protein name	Locus Name	Acc#
	sp:Y746_METJA	Q58156

Description

HYPOTHETICAL PROTEIN MJ0746

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24650302_f1_16	725	5947	723	2172		

Protein name	Locus Name	Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24650912_c3_349.....	726	5948	519	1560	1415	1.0e-144

Protein name	Locus Name	Acc#
sodium/proline symporter (proline permease)	pir:C69115	C69115

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24798457_f3_162.....	727	5949	500	1503		

Protein name	Locus Name	Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24806567_c1_221	728	5950	670	2013	233	1.2e-21

Protein name

Locus Name

Acc#

sp:DSBD_HAEIN

P44919

Description

BIOGENESIS PROTEIN CYCZ)

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24853385_c2_321	729	5951	343	1032		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
2504787_f3_189	730	5952	258	777	89	0.0093

Protein name

Locus Name

Acc#

ORF128 hypothetical protein

gp:AF008210

AF008210

Description

Buchnera aphidicola genomic fragment containing (chaperone Hsp60)groEL, DNA biosynthesis initiating protein (dnaA), ATP operon(atpCDGAHFEB), and putative chromosome replication protein (gidA)genes, complete cds; and termination factor Rho (rho) gene, partialcds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
25859425_c3_373	731	5953	64	195	92	0.00016

Protein name

Locus Name

Acc#

hypothetical protein ssr1765

pir:S74779

S74779

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
26220277_c1_252	732	5954	193	582		

Protein name Locus Name Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
26367943_f2_120	733	5955	341	1026		

Protein name Locus Name Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
26369087_c2_299	734	5956	116	351	263	1.2e-22

Protein name Locus Name Acc#

sp:YHAI_ECOLI P42622

Description

HYPOTHETICAL 13.5 KD PROTEIN IN EXUR-TDCC INTERGENIC REGION

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
26461627_c1_223	735	5957	351	1056	500	4.0e-57

Protein name Locus Name Acc#

sp:HIS8_CANMA P56099

Description

PHOSPHATE TRANSAMINASE)

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
2928387_c1_215	736	5958	197	594		

Protein name Locus Name Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
29314080_c1_231	737	5959	724	2175	1417	6.1e-145

Protein name Locus Name Acc#

sp:DCP_ECOLI

Description

PEPTIDYL-DIPEPTIDASE DCP, (DIPEPTIDYL CARBOXYPEPTIDASE)

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
30256552_f1_40	738	5960	245	738	571	2.7e-55

Protein name Locus Name Acc#

uridine kinase udk

pir:G69728

G69728

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
32228408_f3_176	739	5961	464	1395	473	6.6e-45

Protein name Locus Name Acc#

unknown

gp:AF086638

AF086638

Description

Pseudomonas putida CumA precursor (cumA) and CumB (cumB) genes, complete cds; and unknown genes.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
33235905_c3_385	740	5962	476	1431		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
34183438_f1_35	741	5963	239	720	430	2.4e-40

Protein name

Locus Name

Acc#

sp:YHHW_ECOLI

P46852

Description

HYPOTHETICAL 26.3 KD PROTEIN IN GNTR-GGT INTERGENIC REGION (F231)

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
34625053_f2_129	742	5964	493	1482	352	1.6e-34

Protein name

Locus Name

Acc#

damage-inducible protein PAB0243

pir:A75151

A75151

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
36056510_c1_251	743	5965	422	1269		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
36132912_c2_281	744	5966	167	504	106	0.00018

Protein name

Locus Name

Acc#

hypothetical protein SC2E9.08 SC2E9.08

pir:T34819

T34819

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
3912781_f2_98	745	5967	338	1017	317	1.8e-37

Protein name

Locus Name

Acc#

hypothetical protein F19D11.16:hypothetical protein F14M4.29:hypothetical protein F14M4.29

pir:T02689

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
3928550...f3...161.....	746	5968	1054	3165	325	4.3e-45

Protein name

Locus Name

Acc#

115K outer membrane protein precursor:Susc protein

pir:JC6027

JC6027

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
3938831...f3...147.....	747	5969	416	1251	1961	1.4e-202

Protein name

Locus Name

Acc#

sp:CHUR_BACTN

Q02550

Description

CHONDRO-6-SULFATASE REGULATORY PROTEIN

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
3948562...c1...217.....	748	5970	204	615	373	2.6e-34

Protein name

Locus Name

Acc#

sp:Y120_METTH

O26223

Description

POTATIVE NADH DEHYDROGENASE/NAD(P)H NITROREDUCTASE,

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4022312_c1_209	749	5971	282	849	141	2.7e-09
Protein name			Locus Name		Acc#	
ferredoxin (fdx-3) homolog			pir:C69294		C69294	
Description						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4069137_c3_387	750	5972	301	906	127	1.9e-14
Protein name			Locus Name		Acc#	
leader peptidase Lep			gp:AF188620		AF188620	
Description						
Bordetella pertussis lep operon, complete sequence.						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4117150_f3_197	751	5973	426	1278	1151	9.4e-117
Protein name			Locus Name		Acc#	
			sp:SR54_BACSU		P37105	
Description						
SIGNAL RECOGNITION PARTICLE PROTEIN (FIFTY-FOUR HOMOLOG)						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4461562_c3_370	752	5974	216	651	216	1.1e-17
Protein name			Locus Name		Acc#	
hypothetical protein PAB1763			pir:D75137		D75137	
Description						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4689092_c1_232	753	5975	199	600	244	1.2e-20

Protein name

Locus Name

Acc#

ferric uptake regulator homolog

gp:AF095596

AF095596

Description

Staphylococcus aureus strain ISP3 ferric uptake regulator homolog (furB) gene, complete cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4698432_f2_121	754	5976	340	1023	687	1.4e-67

Protein name

Locus Name

Acc#

synthase III

pir:F70394

F70394

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4734638_c3_354	755	5977	287	864	639	1.7e-62

Protein name

Locus Name

Acc#

sp:HIS1_SALTY

P00499

Description

ATP PHOSPHORIBOSYLTRANSFERASE,

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4876563_c1_216	756	5978	153	462	361	4.9e-33

Protein name

Locus Name

Acc#

sp:SMPB_BACSU

O32230

Description

SMALL PROTEIN B HOMOLOG

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4960812_f3_153	757	5979	158	477	299	1.8e-26

Protein name Locus Name Acc#
sp:THIO_BORBU 051088

Description

THIOREDOXIN (TRX)

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
5175875_c2_320	758	5980	192	579		

Protein name Locus Name Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
5313952_f3_190.....	759	5981	485	1458	1360	6.7e-139

Protein name Locus Name Acc#
raw starch digesting amylase precursor gp:AF067653 AF067653

Description

Cytophaga sp. raw starch digesting amylase precursor, gene, complete cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
579577_f3_152.....	760	5982	158	477	242	2.0e-20

Protein name Locus Name Acc#
thioredoxin-like protein gp:ATAC010718 AC010718

Description

Arabidopsis thaliana chromosome I BAC F28016 genomic sequence, complete sequence.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
6056552_c3_386	761	5983	337	1014		

Protein name Locus Name Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
6102312_c1_250	762	5984	349	1050		

Protein name Locus Name Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
6250003_c1_247	763	5985	598	1797	129	4.9e-05

Protein name Locus Name Acc#

conserved hypothetical protein BB0195 pir:C70124 C70124

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
6442137_c3_352	764	5986	103	312	120	1.7e-07

Protein name Locus Name Acc#

sp:YRPX_STRCO P37977

Description

HYPOTHETICAL 11.1 KD PROTEIN IN RPOX 5' REGION

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
829692_c2_303	765	5987	70	213		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
906903_c3_346	766	5988	287	864		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
978387_c1_248	767	5989	130	393		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
9900327_c3_324	768	5990	237	714	149	1.5e-10

Protein name

Locus Name

Acc#

hypothetical protein PH1670

pir:F71047

F71047

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
11721040_f1_42	769	5991	78	237		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
1256885_f3_133	770	5992	382	1149	510	7.6e-48
<u>Protein name</u>			<u>Locus Name</u>		<u>Acc#</u>	
Man26A			gp:AF126471		AF126471	
<u>Description</u>						
Cellulomonas fimi Man26A (man26A) gene, complete cds.						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
12773255_c2_289	771	5993	519	1560	417	5.7e-39
<u>Protein name</u>			<u>Locus Name</u>		<u>Acc#</u>	
conserved hypothetical protein			pir:B72391		B72391	
<u>Description</u>						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
13709635_c1_211	772	5994	522	1569	319	6.8e-36
<u>Protein name</u>			<u>Locus Name</u>		<u>Acc#</u>	
Arylsulfatase precursor (EC 3.1.6.1)			gp:D90791			
<u>Description</u>						
E.coli genomic DNA, Kohara clone #280(33.7-34.1 min.).						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
13711437_c1_215	773	5995	106	321	148	1.4e-09
<u>Protein name</u>			<u>Locus Name</u>		<u>Acc#</u>	
TRK system potassium uptake protein (trkA)			gp:U32745			
<u>Description</u>						
Haemophilus influenzae Rd section 60 of 163 of the complete genome.						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
14651512_f1_5	774	5996	469	1410	782	1.2e-77

Protein name

Locus Name

Acc#

sp:YAGG_ECOLI

Description

HYPOTHETICAL SYMPORTER IN PERR-ARGF INTERGENIC REGION

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
14726062_c1_203	775	5997	664	1995		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
16600327_f3_130.....	776	5998	216	651	442	1.3e-41

Protein name

Locus Name

Acc#

dimethylamine corrinoid protein MtbC

gp:AF102623

AF102623

Description

Methanosarcina barkeri dimethylamine corrinoid protein MtbC (mtbC), trimethylamine methyltransferase MttB (mttB), trimethylaminecorrinoide protein MttC (mttC), putative transmembrane protein MttP (mttP), and dimethylamine methyltransferase MtbB1 (mtbB1) genes, complete cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
20525252_c2_307.....	777	5999	637	1914	1050	9.7e-125

Protein name

Locus Name

Acc#

sp:NU5C_SYNP2

P31971

Description

NADH-PLASTOQUINONE OXIDOREDUCTASE CHAIN 5,

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

2117177_f2_71	778	6000	439	1320	354	1.5e-31
---------------	-----	------	-----	------	-----	---------

Protein name

Locus Name

Acc#

endo-1,4-beta-mannosidase

pir:D72278

D72278

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

21664650_f1_6	779	6001	396	1191	239	6.0e-18
---------------	-----	------	-----	------	-----	---------

Protein name

Locus Name

Acc#

renin-binding protein-related protein:protein
slr1975:protein slr1975

pir:S75649

S75649

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

21875377_f1_7	780	6002	602	1809		
---------------	-----	------	-----	------	--	--

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

22270002_f2_70	781	6003	258	777	475	4.5e-44
----------------	-----	------	-----	-----	-----	---------

Protein name

Locus Name

Acc#

Man26A

gp:AF126471

AF126471

Description

Cellulomonas fimi Man26A (man26A) gene, complete cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
22550917_c2_317	782	6004	398	1197	99	0.036

Protein name

Locus Name

Acc#

endo-beta-1,3-glucanase precursor

gp:AF013169

Description

Pyrococcus furiosus beta-glucosidase (celB) gene, complete cds; adh-lam operon, complete sequence; biotin ligase BirA homolog(birA) gene, complete cds; and 2-phosphoglycerate kinase (pgk) gene, partial cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
23712837_f2_81	783	6005	377	1134	169	4.8e-12

Protein name

Locus Name

Acc#

conserved hypothetical protein SC9C7.14c

pir:T35965

T35965

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24260302_f3_134	784	6006	398	1197	283	9.0e-25

Protein name

Locus Name

Acc#

conserved hypothetical protein

pir:B72278

B72278

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24409668_f3_124	785	6007	140	423		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24415962_c2_304	786	6008	160	483	231	2.9e-19

Protein name

Locus Name

Acc#

NADH dehydrogenase (ubiquinone), I chain I
RP795

pir:E71640

E71640

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24492137_f1_2	787	6009	1075	3228	163	1.4e-07

Protein name

Locus Name

Acc#

probable secreted glucosidase

pir:T35164

T35164

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24642787_c2_316	788	6010	426	1281		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24644052_c1_205	789	6011	405	1218	197	2.4e-26

Protein name

Locus Name

Acc#

alpha-1,3/4-fucosidase precursor

gp:SSU39394

U39394

Description

Streptomyces sp. alpha-1,3/4-fucosidase precursor gene, completecds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24645437_c3_384	790	6012	901	2706	176	3.8e-08
<u>Protein name</u>			<u>Locus Name</u>		<u>Acc#</u>	
115K outer membrane protein precursor:Susc protein			pir:JC6027		JC6027	
<u>Description</u>						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24648388_f3_198	791	6013	239	717	518	1.1e-49
<u>Protein name</u>			<u>Locus Name</u>		<u>Acc#</u>	
probable glycosyl hydrolase			pir:T36467		T36467	
<u>Description</u>						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
25400260_c1_220	792	6014	279	840	568	5.7e-55
<u>Protein name</u>			<u>Locus Name</u>		<u>Acc#</u>	
			sp:NUOH_ECOLI			
<u>Description</u>						
OXIDOREDUCTASE CHAIN 8) (NUO8)						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
25817656_c3_368	793	6015	836	2511	871	4.4e-87
<u>Protein name</u>			<u>Locus Name</u>		<u>Acc#</u>	
			sp:TRKH_ECOLI			
<u>Description</u>						
TRK SYSTEM POTASSIUM UPTAKE PROTEIN TRKH						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
26230265_f2_67	794	6016	625	1878		

Protein name	Locus Name	Acc#
--------------	------------	------

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
26360717_f3_183	795	6017	219	660	346	1.9e-31

Protein name	Locus Name	Acc#
--------------	------------	------

phosphoglycolate phosphatase (gph) homolog	pir:C70184	C70184
--	------------	--------

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
26366542_c2_308	796	6018	498	1497	719	1.9e-73

Protein name	Locus Name	Acc#
--------------	------------	------

NADH dehydrogenase (ubiquinone), chain 4.2:protein slr1291:protein slr1291	pir:S74687	S74687
--	------------	--------

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
26384637_c3_372	797	6019	531	1596	738	4.7e-85

Protein name	Locus Name	Acc#
--------------	------------	------

NADH dehydrogenase (ubiquinone), I chain nuoD2	pir:D70413	D70413
--	------------	--------

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
26587708_f2_65	798	6020	536	1611	204	3.8e-13

Protein name	Locus Name	Acc#
unknown	gp:U96771	U96771

Description

Prevotella bryantii putative polygalacturonase, B-1,4-endoglucanase, and mannanase genes, complete cds; and unknown genes.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
26594137_f2_75	799	6021	336	1011	293	7.9e-26

Protein name	Locus Name	Acc#
methylcobamide:CoM methyltransferase isozyme A	gp:AF013713	AF013713

Description

Methanosarcina barkeri methylcobamide:CoM methyltransferase isozyme A (mtbA), monomethylamine corrinoid protein (mtmC), monomethylaminemethyltransferase (mtmB), putative monomethylamine permease (mtmP), and unknown genes, complete cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
26604712_c2_309.....	800	6022	485	1458	725	1.3e-71

Protein name	Locus Name	Acc#
	sp:NU2C_SYNY3	P72714

Description

NADH-PLASTOQUINONE OXIDOREDUCTASE CHAIN 2,

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
29457557_c2_300.....	801	6023	126	381	225	1.3e-18

Protein name	Locus Name	Acc#
	sp:NU3C_ANTFO	Q31792

Description

NADH-PLASTOQUINONE OXIDOREDUCTASE CHAIN 3, CHLOROPLAST,

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
31776708_c3_371	802	6024	65	198	163	4.7e-12
<u>Protein name</u>			<u>Locus Name</u>			<u>Acc#</u>
NADH dehydrogenase (ubiquinone), I chain nuoB			pir:C70413			C70413
<u>Description</u>						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
32532838_f2_80	803	6025	518	1557		
<u>Protein name</u>			<u>Locus Name</u>			<u>Acc#</u>
<u>Description</u>						

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
33835790_f2_74	804	6026	251	756	92	0.045
<u>Protein name</u>			<u>Locus Name</u>			<u>Acc#</u>
hypothetical protein			pir:C72397			C72397
<u>Description</u>						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
36132686_c3_364	805	6027	114	345		
<u>Protein name</u>			<u>Locus Name</u>			<u>Acc#</u>
<u>Description</u>						

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
36360962_c2_305	806	6028	172	519	204	2.1e-16
<u>Protein name</u>			<u>Locus Name</u>			<u>Acc#</u>
NADH dehydrogenase (ubiquinone), I chain J			pir:C71839			C71839
<u>Description</u>						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
3907132_f2_90	807	6029	932	2799	470	7.6e-41
Protein name			Locus Name		Acc#	
sensory transduction histidine kinase slr2098:protein slr2098:protein slr2098			pir:S75130		S75130	
Description						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
3995302_c2_301	808	6030	185	558	319	1.4e-28
Protein name			Locus Name		Acc#	
NADH dehydrogenase I, subunit nuoB			gp:ECNU00		X68301	
Description						
E.coli DNA sequence of nuo operon.						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4146907_c1_202	809	6031	1071	3216	714	2.4e-81
Protein name			Locus Name		Acc#	
receptor antigen (RagA)			gp:PGI130872		AJ130872	
Description						
Porphyromonas gingivalis W50 receptor antigen (rag) locus encoding a major immunodominant 55kDa antigen.						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4553400_c2_287	810	6032	694	2085	173	6.3e-18
Protein name			Locus Name		Acc#	
Sip1 protein			pir:S27762		S27762	
Description						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4566876_c2_285	811	6033	487	1464	411	2.5e-38

Protein name

Locus Name

Acc#

sp:YIDJ_ECOLI

P31447

Description

HYPOTHETICAL 57.3 KD PROTEIN IN EMRD-GLVG INTERGENIC REGION

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4975313_c3_369	812	6034	128	387		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
5111038_c2_306	813	6035	105	318	231	2.9e-19

Protein name

Locus Name

Acc#

sp:NULC_PLEBO

Q00244

Description

NADH-PLASTOQUINONE OXIDOREDUCTASE CHAIN 4L,

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
5860027_f3_196	814	6036	1380	4143	519	4.9e-46

Protein name

Locus Name

Acc#

utilizing regulatory protein tutC

gp:TTU57900

U57900

Description

Thauera aromatica utilizing regulatory protein tutC (tutC),utilizing regulatory protein tutB (tutB), putative DNA bindingprotein TutB1 (tutB1), and putative protein kinase TutC1 (tutC1)genes, complete cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
6444137_f3_129	815	6037	421	1266	138	2.8e-06

Protein name: CmuC protein

Locus Name: gp:MSP011317

Acc#: AJ011317

Description: Methylobacterium sp. CM4, cobD, metF, cmuB, cmuC, partial cobC and cobQ, genes and genes encoding Orf219 and Orf361.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
7074155_f1_1	816	6038	384	1155	211	1.4e-14

Protein name: unknown

Locus Name: gp:U96771

Acc#: U96771

Description: Prevotella bryantii putative polygalacturonase, B-1,4-endoglucanase, and mannanase genes, complete cds; and unknown genes.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
7220153_c2_291.....	817	6039	686	2061	1366	1.6e-139

Protein name:

Locus Name: sp:DXS_HAEIN

Acc#: P45205

Description: 1-DEOXYXYLULOSE-5-PHOSPHATE SYNTHASE (DXP SYNTHASE)

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
972167_c1_199.....	818	6040	512	1539	359	3.0e-43

Protein name:

Locus Name: sp:EXUT_ECOLI

Acc#: P42609

Description: HEXURONATE TRANSPORTER

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
9954806_f1_3	819	6041	820	2463	1419	3.8e-145
Protein name			Locus Name		Acc#	
beta-xylo-glucosidase			gp:TBZ56279		Z56279	
Description						
T.brockii cglF, cglG, xglS and cglT genes.						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
33212528_c3_11	820	6042	554	1662		
Protein name			Locus Name		Acc#	
Description						
NO-HIT						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
7119532_f3_6	821	6043	71	216	53	0.017
Protein name			Locus Name		Acc#	
			sp:GP38_CANFA		Q95152	
Description						
GLYCOPROTEIN 38 PRECURSOR (GP38) (MUCIN-TYPE MEMBRANE PROTEIN GP40)						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
10007303_f1_21	822	6044	177	534		
Protein name			Locus Name		Acc#	
Description						
NO-HIT						

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
10189501_f3_191	823	6045	81	246		

Protein name Locus Name Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
10314088_f1_46	824	6046	384	1155	631	1.2e-61

Protein name Locus Name Acc#

sp:DINP_ECOLI

Description

DNA-DAMAGE-INDUCIBLE PROTEIN P

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
10548816_f1_86	825	6047	62	189	102	1.4e-05

Protein name Locus Name Acc#

hypothetical protein APE2457 pir:H72476 H72476

Description

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
1072177_g2_422	826	6048	1110	3333	671	3.2e-184

Protein name Locus Name Acc#

sp:SECA_RHOCA P52966

Description

PREPROTEIN TRANSLOCASE SECA SUBUNIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
11147938_f2_90	827	6049	401	1206	295	9.2e-25

Protein name

Locus Name

Acc#

115K outer membrane protein precursor:Susc protein

pir:JC6027

JC6027

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
1173557_c1_261	828	6050	349	1050	483	5.8e-46

Protein name

Locus Name

Acc#

sp:APBE_HAEIN

P44550

Description

THIAMINE BIOSYNTHESIS LIPOPROTEIN APBE PRECURSOR

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
1180302_f2_93	829	6051	542	1629	256	1.3e-40

Protein name

Locus Name

Acc#

sp:STS_RAT

P15589

Description

SULFATE SULFOHYDROLASE) (ARYLSULFATASE C) (ASC)

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
1208392_c2_362	830	6052	541	1626	396	9.6e-37

Protein name

Locus Name

Acc#

sp:RLUA_ECOLI

P39219

Description

(PSEUDOURIDYLATE SYNTHASE) (URACIL HYDROLYASE)

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
128775_f1_51	831	6053	166	501		

<u>Protein name</u>	<u>Locus Name</u>	<u>Acc#</u>

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
12897563_c3_443	832	6054	307	924	488	1.7e-46

<u>Protein name</u>	<u>Locus Name</u>	<u>Acc#</u>

oxidoreductase, short chain
dehydrogenase/reductase family

pir:E72427 E72427

Description

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
130001_c3_460	833	6055	540	1623	383	2.1e-72

<u>Protein name</u>	<u>Locus Name</u>	<u>Acc#</u>

sp:YFCC_ECOLI

Description

HYPOTHETICAL 54.8 KD PROTEIN IN PTA-FOLX INTERGENIC REGION

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
13881262_f2_105	834	6056	445	1338		

<u>Protein name</u>	<u>Locus Name</u>	<u>Acc#</u>

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
14331502_f3_177	835	6057	166	501		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
14703962_c1_334.....	836	6058	396	1191		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
14723751_c2_360.....	837	6059	673	2022	778	1.1e-79

Protein name

Locus Name

Acc#

type III DNA modification enzyme
(methyltransferase)

pir:F71810

F71810

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
14876578_t1_37.....	838	6060	296	891	310	1.2e-27

Protein name

Locus Name

Acc#

probable beta-glycosyltransferase trsC

pir:S51262

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
15054691_c1_330.....	839	6061	430	1293		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
156642_c3_469	840	6062	133	402		

Protein name Locus Name Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
15902_f2_153	841	6063	264	795	316	2.9e-28

Protein name Locus Name Acc#

sp:YDAO_ECOLI

Description

HYPOTHETICAL 35.6 KD PROTEIN IN DBPA-INTR INTERGENIC REGION

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
163515_c3_466	842	6064	62	189		

Protein name Locus Name Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
16828575_f1_43	843	6065	388	1167	790	1.6e-80

Protein name Locus Name Acc#

GTP-binding protein

gp:AF019407

AF019407

Description

Caulobacter crescentus GTP-binding protein (cgTA) gene, complete cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
16829461_f2_112	844	6066	119	360	108	3.2e-06
<u>Protein name</u>				<u>Locus Name</u>		<u>Acc#</u>
hypothetical protein PH0360				pir:E71143		E71143
<u>Description</u>						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
16832885_f2_159	845	6067	431	1296	1723	2.3e-177
<u>Protein name</u>				<u>Locus Name</u>		<u>Acc#</u>
hypothetical protein				pir:JQ1020		JQ1020
<u>Description</u>						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
19703461_c1_311	846	6068	481	1446	1581	2.6e-162
<u>Protein name</u>				<u>Locus Name</u>		<u>Acc#</u>
unknown				gp:AF048749		AF048749
<u>Description</u>						
Bacteroides fragilis capsular polysaccharide biosynthesis operon, complete sequence.						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
19725250_c2_343	847	6069	202	609	327	2.0e-29
<u>Protein name</u>				<u>Locus Name</u>		<u>Acc#</u>
				sp:Y112_METTH		027840
<u>Description</u>						
HYPOTHETICAL PROTEIN MTH1812						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
19797162_f2_119	848	6070	357	1074	1095	8.1e-111
<u>Protein name</u>		<u>Locus Name</u>		<u>Acc#</u>		
nucleotide sugar epimerase		gp:AF059755		AF059755		
<u>Description</u>						
Vibrio vulnificus nucleotide sugar epimerase gene, complete cds.						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
1992187_f3_212	849	6071	355	1068	193	8.7e-15
<u>Protein name</u>		<u>Locus Name</u>		<u>Acc#</u>		
lumQ protein:protein slr1213:protein slr1213		pir:S77548		S77548		
<u>Description</u>						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
20050402_f3_253.....	850	6072	163	492	129	1.8e-07
<u>Protein name</u>		<u>Locus Name</u>		<u>Acc#</u>		
phosphopyruvate hydratase		pir:C75251		C75251		
<u>Description</u>						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
20087751_f3_199.....	851	6073	319	960	1657	2.3e-170
<u>Protein name</u>		<u>Locus Name</u>		<u>Acc#</u>		
putative UDP-GlcNAc:undecaprenylphosphate		gp:AF048749		AF048749		
<u>Description</u>						
Bacteroides fragilis capsular polysaccharide biosynthesis operon, complete sequence.						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
2037502_f1_17	852	6074	256	771	245	9.6e-21
Protein name			Locus Name		Acc#	
conserved hypothetical protein			pir:D72320		D72320	
Description						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
20734625_f3_208	853	6075	223	672	225	1.3e-18
Protein name			Locus Name		Acc#	
hypothetical protein			gp:SSU18930		Y18930	
Description						
Sulfolobus solfataricus 281 kb genomic DNA fragment, strain P2.						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
20754427_f3_198	854	6076	347	1044	1691	5.7e-174
Protein name			Locus Name		Acc#	
UDP-glucose-4-epimerase/dTDP-glucose-4,6			gp:AF048749		AF048749	
Description						
Bacteroides fragilis capsular polysaccharide biosynthesis operon, complete sequence.						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
211510_f1_57	855	6077	342	1029	307	2.6e-27
Protein name			Locus Name		Acc#	
activator protein			gp:AF047527		AF047527	
Description						
Pseudomonas fluorescens activator protein (mtIR) gene, complete cds.						

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
21640887_f2_117	856	6078	362	1089	157	1.7e-08
<u>Protein name</u>			<u>Locus Name</u>		<u>Acc#</u>	
hypothetical protein 7.17			pir:D47677			
<u>Description</u>						

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
21681552_g3_442	857	6079	466	1401	793	8.2e-79
<u>Protein name</u>			<u>Locus Name</u>		<u>Acc#</u>	
thiophene and furan oxidation protein			pir:C70375		C70375	
<u>Description</u>						

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
22536527_f3_190	858	6080	215	648	1136	3.7e-115
<u>Protein name</u>			<u>Locus Name</u>		<u>Acc#</u>	
putative methyl transferase			gp:AF048749		AF048749	
<u>Description</u>						
Bacteroides fragilis capsular polysaccharide biosynthesis operon, complete sequence.						

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
22540937_f1_4	859	6081	522	1569	304	1.1e-45
<u>Protein name</u>			<u>Locus Name</u>		<u>Acc#</u>	
			sp:STS_HUMAN		P08842	
<u>Description</u>						
SULFATE SULFOHYDROLASE) (ARYLSULFATASE C) (ASC)						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
22860128_f3_255	860	6082	83	252	64	0.031

Protein name

Locus Name

Acc#

sp:SPRC_XENLA

P36378

Description

(OSTEONECTIN) (ON) (BASEMENT MEMBRANE PROTEIN BM-40)

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
23472533_f3_254	861	6083	124	375	132	8.7e-08

Protein name

Locus Name

Acc#

phosphopyruvate hydratase

pir:C75251

C75251

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
23472066_c3_434	862	6084	183	552		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
23647758_c2_399	863	6085	81	246		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
23709625_c3_433	864	6086	321	966	627	3.2e-61

Protein name

Locus Name

Acc#

dolichol-phosphate mannosyltransferase

pir:G70463

G70463

Description

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
24064142_f2_148	865	6087	265	798	362	3.8e-33
<u>Protein name</u>	<u>Locus Name</u>				<u>Acc#</u>	
hypothetical protein ywnB	pir:E70063				E70063	
<u>Description</u>						

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
24114142_c2_373	866	6088	291	876		
<u>Protein name</u>	<u>Locus Name</u>				<u>Acc#</u>	
<u>Description</u>						

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
24229677_c2_340	867	6089	238	717	280	1.9e-24
<u>Protein name</u>	<u>Locus Name</u>				<u>Acc#</u>	
hypothetical protein yisX	pir:G69838				G69838	
<u>Description</u>						

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
24245437_c2_423	868	6090	153	462		
<u>Protein name</u>	<u>Locus Name</u>				<u>Acc#</u>	
<u>Description</u>						
NO-HIT						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24257187_f2_132	869	6091	411	1236	1104	9.0e-112

Protein name	Locus Name	Acc#
putative carboxybiotin decarboxylase subunit of	gp:MRU87980	U87980

Description

Malonomonas rubra putative IS-element gene, partial cds, andmalonate decarboxylase gene cluster (madY, madZ, madG, madB, madA, madE, madC, madD, madH, madK, madF, madL, madM, madN) genes, complete cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24401507_c1_299	870	6092	510	1533	2702	4.2e-281

Protein name	Locus Name	Acc#
unknown	gp:AF048749	AF048749

Description

Bacteroides fragilis capsular polysaccharide biosynthesis operon, complete sequence.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24473192_c2_372	871	6093	642	1929	110	0.0037

Protein name	Locus Name	Acc#
	sp:YOBW_MYCLE	Q49757

Description

HYPOTHETICAL 31.1 KD PROTEIN B1937_F2_39

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24473817_f2_106	872	6094	431	1296	457	3.3e-43

Protein name	Locus Name	Acc#
putative hemolysin	gp:AF051356	AF051356

Description

Streptococcus mutans YtqB (ytqB) gene, partial cds; ABC transporter(abcX), putative permease (perM), putative hemolysin (hlyX), pyruvate-formate lyase activating enzyme (pflC), D-alanine-D-alanyl carrier protein ligase (dltA), integral membrane protein (dltB), D-alanyl carrier protein (dltC), extramembranal protein (dltD), and putative exopolyphosphatase (ppx1) genes,

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24495337_c2_345.....	873	6095	499	1500	119	0.00014

Protein name	Locus Name	Acc#
immunogenic 75 kDa protein PG4	gp:AF145800	AF145800

Description

Porphyromonas gingivalis strain W50 immunogenic 75 kDa protein PG4 gene, complete cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24631305_f2_116.....	874	6096	183	552	575	1.0e-55

Protein name	Locus Name	Acc#
unknown	gp:AF048749	AF048749

Description

Bacteroides fragilis capsular polysaccharide biosynthesis operon, complete sequence.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24647763_c1_275.....	875	6097	498	1497	685	2.3e-67

Protein name	Locus Name	Acc#
	sp:RIBB_ECOLI	P24199

Description

3,4-DIHYDROXY-2-BUTANONE 4-PHOSPHATE SYNTHASE (DHBP SYNTHASE)

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24651515_f2_91	876	6098	108	327		

Protein name	Locus Name	Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24666000_c3_437	877	6099	306	921	150	4.1e-16

Protein name	Locus Name	Acc#

probable uridine phosphorylase APE2105	pir:D72516	D72516
--	------------	--------

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24804663_f3_195	878	6100	386	1161		

Protein name	Locus Name	Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
25972937_c3_459	879	6101	69	210		

Protein name	Locus Name	Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
26212777_c3_450	880	6102	300	903	215	1.4e-17

Protein name	Locus Name	Acc#

hypothetical protein sll1671	pir:S74655	S74655
------------------------------	------------	--------

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
26261313_f1_58	881	6103	63	192		

Protein name Locus Name Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
26444687_c2_341	882	6104	434	1305	1588	4.6e-163

Protein name Locus Name Acc#

sp:ENO_STAAU 069174

Description

GLYCERATE HYDRO-LYASE) (LAMININ BINDING PROTEIN)

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
26594206_f1_42	883	6105	180	543	896	9.9e-90

Protein name Locus Name Acc#

putative hypoxanthine guanine gp:AF048749 AF048749

Description

Bacteroides fragilis capsular polysaccharide biosynthesis operon, complete sequence.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
26594686_c2_346	884	6106	158	477		

Protein name Locus Name Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
26595337_f3_180	885	6107	175	528		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
26664012_c2_404	886	6108	355	1068	112	0.0032

Protein name

Locus Name

Acc#

gamma response I protein

gp:ATH131708

AJ131708

Description

Arabidopsis thaliana gr I gene, exons 1-3.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
26761057_c2_361	887	6109	1017	3054	627	1.0e-119

Protein name

Locus Name

Acc#

restriction endonuclease

gp:AF060119

AF060119

Description

Pasteurella haemolytica methyltransferase (mod) and restriction endonuclease (res) genes, complete cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
26836680_f2_109	888	6110	416	1251	1189	8.9e-121

Protein name

Locus Name

Acc#

immunoreactive 47 kD antigen PG120

gp:AF144640

AF144640

Description

Porphyromonas gingivalis strain W50 immunoreactive 47 kD antigen PG120 gene, complete cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
275125_f2_111	889	6111	469	1410	111	2.7e-06

Protein name	Locus Name	Acc#
hypothetical protein Rv2333c	pir:F70705	F70705
<u>Description</u>		

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
2812827_f3_207	890	6112	270	813	298	2.3e-26

Protein name	Locus Name	Acc#
	sp:YFIH_HAEIN	P44552
<u>Description</u>		

HYPOTHETICAL PROTEIN HI0175

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
29336040_f1_36	891	6113	420	1263	144	7.9e-07

Protein name	Locus Name	Acc#
NADH dehydrogenase (ubiquinone), chain 2	pir:T11319	T11319
<u>Description</u>		

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
29703165_f2_118	892	6114	396	1191	369	6.9e-34

Protein name	Locus Name	Acc#
	sp:CAPA_BACAN	P19579
<u>Description</u>		

CAPA PROTEIN

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
30084688_f2_127	893	6115	60	183		

Protein name Locus Name Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
3023452_c3_522	894	6116	524	1575	572	2.1e-55

Protein name Locus Name Acc#

alkaline phosphatase gp:SSPPHOA2 Z48801

Description

Synechococcus PCC7942 phoV gene for alkaline phosphatase.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
31330037_c2_347	895	6117	182	549	204	6.9e-15

Protein name Locus Name Acc#

DNA polymerase III, alpha subunit pir:C72360 C72360

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
3240675_f3_182	896	6118	135	408	115	7.2e-07

Protein name Locus Name Acc#

protein-export membrane protein pir:E71837 E71837

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
33301250_c2_380	897	6119	421	1266		

Protein name Locus Name Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
33397811_f1_45	898	6120	898	2697	115	0.00042

Protein name	Locus Name	Acc#
histidine kinase sensor protein	pir:D70328	D70328

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
33489041_f2_110	899	6121	260	783	94	0.045

Protein name	Locus Name	Acc#
	sp:TPMN_XENLA	Q01174

Description

TROPOMYOSIN ALPHA CHAIN, NON MUSCLE

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
33881887_f3_192	900	6122	83	252	69	0.042

Protein name	Locus Name	Acc#
	sp:YA49_HAEIN	

Description

HYPOTHETICAL PROTEIN HI1049

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
34189385_f1_38	901	6123	311	936	514	3.0e-49

Protein name	Locus Name	Acc#
	gp:BCY11138	Y11138

Description

B.cereus DNA for ORF1, ORF2 and ORF3 (2402 bp).

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
34407193_f1_47	902	6124	144	435		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
34412513_f3_194	903	6125	303	912	193	1.4e-14

Protein name

Locus Name

Acc#

glycosyl transferase PAB0772

pir:B75096

B75096

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
34640915_f1_53	904	6126	63	192		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
34650341_f3_230	905	6127	79	240		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
34652167_f2_108	906	6128	375	1128	530	6.0e-51

Protein name

Locus Name

Acc#

pyridoxal phosphate biosynthetic protein PdxA

pir:H70373

H70373

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
34666452_f3_224	907	6129	172	519		

Protein name	Locus Name	Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
35134637_f3_249	908	6130	123	372	221	3.3e-18

Protein name	Locus Name	Acc#

hypothetical protein	pir:H75473	H75473
----------------------	------------	--------

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
3944005_f1_39	909	6131	440	1323	1205	1.8e-122

Protein name	Locus Name	Acc#

putative UDP-glucose dehydrogenase	gp:AF159428	AF159428
------------------------------------	-------------	----------

Description

Burkholderia pseudomallei putative UDP-glucose dehydrogenase (udg), putative ADP-heptose synthase (waaE), and putative ADP-glycero-mannoheptose epimerase (gmhD) genes, complete cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
3988318_c2_384	910	6132	699	2100	3614	0.0

Protein name	Locus Name	Acc#

putative TonB-dependent outer membrane receptor	gp:AF048749	AF048749
---	-------------	----------

Description

Bacteroides fragilis capsular polysaccharide biosynthesis operon, complete sequence.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
3990900_f3_193	911	6133	392	1179	272	1.3e-31
Protein name			Locus Name		Acc#	
probable galactosyltransferase trsd			pir:S51263			
Description						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4065757_c1_259	912	6134	262	789		
Protein name			Locus Name		Acc#	
Description						

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4094512_c3_436	913	6135	344	1035	1054	1.8e-106
Protein name			Locus Name		Acc#	
			sp:YQFA_BACSU		P54466	
Description						

HYPOTHETICAL 35.6 KD PROTEIN IN RPSU-PHOH INTEREGENIC REGION

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4116680_f3_175	914	6136	245	738		
Protein name			Locus Name		Acc#	
Description						

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4119677_f2_125	915	6137	248	747	198	1.2e-15
Protein name			Locus Name		Acc#	
hypothetical protein jhp1456			pir:C71806		C71806	
Description						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4147280_c2_374	916	6138	67	204		

Protein name	Locus Name	Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
425010_f3_196	917	6139	385	1158	642	8.2e-63

Protein name	Locus Name	Acc#

WbpU

gp:AF035937

AF035937

Description

Pseudomonas aeruginosa strain IATS O6 RpsA (rpsA) gene, partialcds; Ihf-Beta, Wzz (wzz), and Wzx (wzx) genes, complete cds; andwbp gene cluster for O-antigen biosynthesis, complete sequence.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
427215_f3_162	918	6140	473	1422	143	1.6e-06

Protein name	Locus Name	Acc#

unknown

gp:U96771

U96771

Description

Prevotella bryantii putative polygalacturonase, B-1,4-endoglucanase, and mannanase genes, complete cds; and unknowngenes.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4323262_c1_273	919	6141	955	2868		

Protein name	Locus Name	Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4331300_c2_390	920	6142	82	249	103	1.1e-05
Protein name			Locus Name		Acc#	
hypothetical protein			gp:SSU18930		Y18930	
Description						
Sulfolobus solfataricus 281 kb genomic DNA fragment, strain P2.						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4409462_c2_403	921	6143	506	1521	631	1.2e-61
Protein name			Locus Name		Acc#	
conserved hypothetical protein aq_1365			pir:F70418		F70418	
Description						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4694152_f2_107	922	6144	717	2154	122	0.00078
Protein name			Locus Name		Acc#	
putative peptidyl-prolyl cis-trans isomerase			gp:ASAJ2316		AJ002316	
Description						
Acinetobacter sp. ADP1 alkR & alkM genes, ORF1 & ORF4.						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4725257_f2_104	923	6145	427	1284	89	0.011
Protein name			Locus Name		Acc#	
membrane protein			gp:PPUY18245		Y18245	
Description						
Pseudomonas putida todX, todF, todC1, todC2, todB, todA, todD, todE, todG, todI, todH, todS, todT genes.						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4804632_c3_476	924	6146	218	657	1119	2.3e-113

Protein name	Locus Name	Acc#
unknown	gp:AF048749	AF048749

Description

Bacteroides fragilis capsular polysaccharide biosynthesis operon, complete sequence.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
5110700_f1_35	925	6147	485	1458	780	1.9e-77

Protein name	Locus Name	Acc#
O-antigen repeat unit transporter Wzx	gp:AF172324	AF172324

Description

Escherichia coli GalF (galF) gene, partial cds; O-antigen repeatunit transporter Wzx (wzx), WbnA (wbnA), O-antigen polymerase Wzy (wzy), WbnB (wbnB), WbnC (wbnC), WbnD (wbnD), WbnE (wbnE), UDP-Glc-4-epimerase GalE (galE), 6-phosphogluconate dehydrogenase Gnd (gnd), UDP-Glc-6-dehydrogenase Ugd (ugd), and WbnF (wbnF) genes, complete cds; and chain length determinant

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
5270252_f2_124	926	6148	192	579	335	3.4e-40

Protein name	Locus Name	Acc#
	gp:AB017508	AB017508

Description

Bacillus halodurans C-125 genomic DNA, 32 kb fragment, complete cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
5313775_f1_34	927	6149	157	474	614	7.6e-60

Protein name	Locus Name	Acc#
unknown	gp:AF048749	AF048749

Description

Bacteroides fragilis capsular polysaccharide biosynthesis operon, complete sequence.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
553212_f3_241	928	6150	654	1965	485	6.4e-45
<u>Protein name</u>				<u>Locus Name</u>	<u>Acc#</u>	
sensory transduction histidine kinase slr2098:protein slr2098:protein slr2098				pir:S75130	S75130	
<u>Description</u>						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
6.57956_c1_258	929	6151	210	633		
<u>Protein name</u>				<u>Locus Name</u>	<u>Acc#</u>	
<u>Description</u>						

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
6.671887_f3_238	930	6152	1294	3885	384	6.5e-33
<u>Protein name</u>				<u>Locus Name</u>	<u>Acc#</u>	
putative alpha-glucosidase				gp:AAC252161	AJ252161	
<u>Description</u>						

Alicyclobacillus acidocaldarius maltose/maltodextrine transportgene region
(maleFGR genes, cdaA gene and glcA gene).

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
6.687552_f1_19	931	6153	349	1050	602	1.4e-58
<u>Protein name</u>				<u>Locus Name</u>	<u>Acc#</u>	
				sp:YFGB_ECOLI	P36979	
<u>Description</u>						

HYPOTHETICAL 43.1 KD PROTEIN IN NDK-GCPE INTERGENIC REGION

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
6767537_f1_13	932	6154	75	228		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
6854762_c3_470	933	6155	68	207		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
7225916_f1_83	934	6156	65	198		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
796875_f2_158	935	6157	60	183		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
961067_c3_479	936	6158	63	192		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
9813_c3_473	937	6159	65	198		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
9921927_c1_288.....	938	6160	419	1260	568	5.7e-55

Protein name

Locus Name

Acc#

conserved hypothetical protein ykgB

pir:D69856

D69856

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
10009632_c1_84.....	939	6161	598	1797	75	0.032

Protein name

Locus Name

Acc#

unknown

gp:U96771

U96771

Description

Prevotella bryantii putative polygalacturonase, B-1,4-endoglucanase, and mannanase genes, complete cds; and unknown genes.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
10753760_f2_32.....	940	6162	131	396	224	1.6e-18

Protein name

Locus Name

Acc#

IgA Fc receptor-like protein A428L

pir:T17931

T17931

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
12787768_f3_65	941	6163	67	204		

Protein name Locus Name Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
13953388_c1_94	942	6164	296	891	133	2.2e-06

Protein name Locus Name Acc#

sp:VIRF_YEREN P13225

Description

VIRULENCE REGULON TRANSCRIPTIONAL ACTIVATOR

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
14511086_c1_86	943	6165	330	993	212	1.5e-18

Protein name Locus Name Acc#

hypothetical protein F14F9.5 pir:T33774 T33774

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
16832885_f3_68	944	6166	431	1296	1723	2.3e-177

Protein name Locus Name Acc#

hypothetical protein pir:JQ1020 JQ1020

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
22860128_f1_11	945	6167	83	252	64	0.031

Protein name

Locus Name

Acc#

sp:SPRC_XENLA

P36378

Description

(OSTEONECTIN) (ON) (BASEMENT MEMBRANE PROTEIN BM-40)

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
23469691_c3_111	946	6168	179	540		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
23515762_f2_52.....	947	6169	92	279		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
23984402_c1_82.....	948	6170	711	2136	488	8.9e-44

Protein name

Locus Name

Acc#

receptor antigen (RagA)

gp:PGI130872

AJ130872

Description

Porphyromonas gingivalis W50 receptor antigen (rag) locus encoding a major immunodominant 55kDa antigen.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24640675_c2_99	949	6171	511	1536		

Protein name Locus Name Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24806300_f2_40	950	6172	267	804		

Protein name Locus Name Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
25706687_c1_83	951	6173	415	1248	162	2.3e-17

Protein name Locus Name Acc#

receptor antigen (RagA) gp:PGI130872 AJ130872

Description

Porphyromonas gingivalis W50 receptor antigen (rag) locus encoding a major immunodominant 55kDa antigen.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
26360636_f1_5	952	6174	201	606	208	8.0e-17

Protein name Locus Name Acc#

gp:AHU56832 U56832

Description

Aeromonas hydrophila FK506 binding protein (fkpA) gene, completed in 3.9 kb fragment.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
2813912_c3_110	953	6175	460	1383	508	1.3e-48

Protein name
Locus Name
 sp:YHAM_ECOLI

Acc#
 P42626

Description

HYPOTHETICAL 19.4 KD PROTEIN IN EXUR-TDCC INTERGENIC REGION (F188)

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
3417677_c1_85	954	6176	354	1065	171	3.3e-12

Protein name
 KIAA0879 protein

Locus Name
 gp:AB020686

Acc#
 AB020686

Description

Homo sapiens mRNA for KIAA0879 protein, complete cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
3.7925..f1..10.....	955	6177	63	192	54	0.020

Protein name
Locus Name
 gp:AFSCR

Acc#
 X70080

Description

A.franciscana Scr gene (homologue of Drosophila Sex combs reduced).

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4.74167..c3..105.....	956	6178	559	1680		

Protein name
Locus Name
Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4742812_f3_64	957	6179	378	1137	388	7.2e-50
<u>Protein name</u>			<u>Locus Name</u>		<u>Acc#</u>	
hypothetical protein			gp:ATH132745		AJ132745	
<u>Description</u>						
Arabidopsis thaliana hypothetical protein, clone EMG9a29.						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4804562_c2_103	958	6180	452	1359	156	2.6e-11
<u>Protein name</u>			<u>Locus Name</u>		<u>Acc#</u>	
putative outer membrane porin			gp:AF030977			
<u>Description</u>						
Vibrio cholerae glutamyl tRNA synthetase (glTX) gene, partial cds; putative outer membrane porin (ompA), unknown protein, vibriobactin receptor precursor (viuA), and ViuB protein (viuB) genes, complete cds; and VibF (vibF) gene, partial cds.						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4875035_c1_80	959	6181	193	582	180	7.4e-14
<u>Protein name</u>			<u>Locus Name</u>		<u>Acc#</u>	
RNA polymerase sigma factor SigZ-like protein			gp:AF137263		AF137263	
<u>Description</u>						
Bacteroides thetaiotaomicron 30S ribosomal protein S16-like protein, fucose gene cluster, and RNA polymerase sigma factor SigZ-like protein (sigZ) genes, complete cds.						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4882012_c1_95	960	6182	377	1131	253	1.4e-19
<u>Protein name</u>			<u>Locus Name</u>		<u>Acc#</u>	
			gp:AF083424		AF083424	
<u>Description</u>						
Ateline herpesvirus 3 complete genome.						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
5351507_f2_39	961	6183	378	1137		

Protein name Locus Name Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
5881877_c3_107	962	6184	352	1059	147	1.3e-07

Protein name Locus Name Acc#

transmembrane sensor gp:AF051691 AF051691

Description

Pseudomonas aeruginosa stress factor A (psfA), ECF sigma factor (fiuI), transmembrane sensor (fiuR), and hydroxamate-type ferrisiderophore receptor (fiuA) genes, complete cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
970385_c3_108	963	6185	824	2475	209	1.2e-13

Protein name Locus Name Acc#

serine/threonine protein kinase related protein pir:H69064 H69064

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
1064063_c2_139	964	6186	297	894	126	0.00030

Protein name Locus Name Acc#

115K outer membrane protein precursor: SusC protein pir:JC6027 JC6027

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
10742332_c1_106	965	6187	85	258		

Protein name Locus Name Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
11836662_f2_46	966	6188	659	1980	1330	1.0e-135

Protein name Locus Name Acc#

sp:YFIC_BACSU P54719

Description

HYPOTHETICAL ABC TRANSPORTER ATP-BINDING PROTEIN 2 IN GLVBC 3' REGION

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
16284692_c2_135	967	6189	64	195		

Protein name Locus Name Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
20507827_f3_79	968	6190	574	1725	1246	8.1e-127

Protein name Locus Name Acc#

ABC transporter, ATP-binding protein pir:E72396 E72396

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
23562802_c3_144	969	6191	421	1266	575	1.0e-55

Protein name

Locus Name

Acc#

sp:SBCE_RHOCA

068033

Description

EXONUCLEASE SBCE HOMOLOG

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24651557_c2_138	970	6192	380	1143	78	0.038

Protein name

Locus Name

Acc#

Fibronectin type III

gp:HUMFN3A

M12549

Description

Human fibronectin gene type III homology unit corresponding to the cell-binding domain, exons 6 and 7.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24666005_c3_145	971	6193	996	2991	464	1.5e-84

Protein name

Locus Name

Acc#

probable exonuclease,

pir:T03465

T03465

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
25978516_c2_119	972	6194	98	297		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
273442_c2_117	973	6195	363	1092	180	2.7e-11

Protein name	Locus Name	Acc#
cation efflux system (czcB-like)	pir:C70415	C70415
Description		

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
2947138_f1_6	974	6196	195	588		

Protein name	Locus Name	Acc#
Description		

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
29940912_f3_90	975	6197	345	1038	279	2.4e-24

Protein name	Locus Name	Acc#
hypothetical protein TM1693	pir:G72223	G72223
Description		

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
3157813_f1_7	976	6198	390	1173	305	4.2e-27

Protein name	Locus Name	Acc#
probable phosphoesterase, ykuE	pir:B69865	B69865
Description		

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
34173431_f1_5	977	6199	180	543	183	3.6e-14

Protein name

Locus Name

Acc#

SigX

gp:AF115334

Description

Pseudomonas fluorescens PpsA (ppsA) gene, partial cds; EstX (estX), MenG (menG), CmaX (cmaX), CrfX (crfX), CmpX (cmpX), SigX (sigX), OprF (oprF), and CobA (cobA) genes, complete cds; and unknown gene.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
34661301_c1_102	978	6200	1083	3252	394	6.3e-53

Protein name

Locus Name

Acc#

acriflavine resistance protein (acrB) homolog

pir:D70117

D70117

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
3938215_c2_118	979	6201	550	1653	384	7.4e-33

Protein name

Locus Name

Acc#

cation efflux (AcrB/AcrD/AcrF family)

pir:F70368

F70368

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4394642_f3_71	980	6202	152	459		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
4805286_c1_99	981	6203	486	1461	533	2.7e-50
<u>Protein name</u>	<u>Locus Name</u>				<u>Acc#</u>	
acriflavine resistance protein (acrB) homolog	pir:D70117				D70117	
<u>Description</u>						

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT</u> <u>Length</u>	<u>AA</u> <u>Length</u>	<u>Score</u>	<u>Probability</u>
5195317_c1_101.....	982	6204	430	1293	110	0.0047
<u>Protein name</u>				<u>Locus Name</u>		<u>Acc#</u>
				sp:YD40_HAEIN		P44165
<u>Description</u>						
HYPOTHETICAL PROTEIN HI1340						

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT</u> <u>Length</u>	<u>AA</u> <u>Length</u>	<u>Score</u>	<u>Probability</u>
6853436_c1_115.....	983	6205	161	486		
<u>Protein name</u>				<u>Locus Name</u>		<u>Acc#</u>
<u>Description</u>						
NO-HIT						

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT</u> <u>Length</u>	<u>AA</u> <u>Length</u>	<u>Score</u>	<u>Probability</u>
10235877_c2_250.....	984	6206	389	1170	2007	1.8e-207
<u>Protein name</u>	<u>Locus Name</u>				<u>Acc#</u>	
putative epimerase/dehydratase				gp:AF125164		AF125164
<u>Description</u>						

Bacteroides fragilis 638R polysaccharide B (PS B2) biosynthesis locus, complete sequence; and unknown genes.

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
10585817_f3_105	985	6207	66	201		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
1064005_f1_40	986	6208	632	1899	291	5.0e-23

Protein name

Locus Name

Acc#

hypothetical protein Rv2731

pir:B70506

B70506

Description

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
10667943_f2_70	987	6209	113	342	125	5.0e-08

Protein name

Locus Name

Acc#

HipA protein.

gp:D90794

Description

E.coli genomic DNA, Kohara clone #303(34.3-34.6 min.).

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
10969017_c3_293	988	6210	343	1032	1742	2.2e-179

Protein name

Locus Name

Acc#

putative epimerase/dehydratase

gp:AF125164

AF125164

Description

Bacteroides fragilis 638R polysaccharide B (PS B2) biosynthesis locus, complete sequence; and unknown genes.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
11023432_c1_205	989	6211	414	1245	2081	2.7e-215

Protein name	Locus Name	Acc#
putative glycosyltransferase	gp:AF125164	AF125164

Description

Bacteroides fragilis 638R polysaccharide B (PS B2) biosynthesis locus, complete sequence; and unknown genes.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
1188951_f2_61	990	6212	60	183		

Protein name	Locus Name	Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
12187817_f3_117.....	991	6213	162	489	93	0.031

Protein name	Locus Name	Acc#
cell cycle progression restoration 8 protein	gp:AF011794	AF011794

Description

Homo sapiens cell cycle progression restoration 8 protein (CPR8)mRNA, complete cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
12520688_c3_280.....	992	6214	61	186		

Protein name	Locus Name	Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
13804187_f1_47	993	6215	98	297	84	0.0018

Protein name	Locus Name	Acc#
hypothetical protein	gp:MTH243656	AJ243656

Description

Methanobacterium thermoautotrophicum ehbA, B, C, D, E, F, G, H, I, J, K, L, M, N, O, P, Q, & ORFS 1, 2 & 3.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
14250637_f3_147	994	6216	387	1164		

Protein name	Locus Name	Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
14258450_f3_118.....	995	6217	121	366	100	2.2e-05

Protein name	Locus Name	Acc#
hypothetical protein TM1330	pir:F72267	F72267

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
14275252_f3_162.....	996	6218	681	2046	1133	7.6e-115

Protein name	Locus Name	Acc#
(p)ppGpp synthetase	gp:BSU86377	U86377

Description

Bacillus subtilis (p)ppGpp synthetase (relA) and adeninephosphoribosyltransferase (apt) genes, complete cds.

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
14648380_f1_18	997	6219	295	888		

Protein name Locus Name Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
14650882_f3_114	998	6220	121	366	96	0.00067

Protein name Locus Name Acc#

hypothetical protein PFB0225c pir:E71620 E71620

Description

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
14660952_f3_150	999	6221	119	360	220	4.3e-18

Protein name Locus Name Acc#

ybeB protein homolog iojap:protein slr1886:protein slr1886 pir:S77145 S77145

Description

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
14925162_f3_119	1000	6222	74	225		

Protein name Locus Name Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
15709675_c1_193	1001	6223	367	1104	798	2.4e-79
Protein name			Locus Name		Acc#	
			sp:YS18_MYCTU		P71777	

Description

HYPOTHETICAL 36.3 KD PROTEIN CY277.18

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
15790675_c1_199	1002	6224	401	1206	862	4.0e-86
Protein name			Locus Name		Acc#	
phosphonopyruvate decarboxylase, fom2			pir:S60212		S60212	

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
19533462_c3_278	1003	6225	419	1260	869	7.2e-87
Protein name			Locus Name		Acc#	
			sp:YBDG_ECOLI			

Description

HYPOTHETICAL 46.6 KD PROTEIN IN PHEP-NFNB INTERGENIC REGION

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
1964058_f2_92	1004	6226	279	840	716	1.2e-70
Protein name			Locus Name		Acc#	
			sp:SOJ_BACSU		P37522	

Description

SOJ PROTEIN

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
20736678_c3_296	1005	6227	204	615	1056	1.1e-106

Protein name	Locus Name	Acc#
putative undecaprenyl-phosphate	gp:AF125164	AF125164

Description

Bacteroides fragilis 638R polysaccharide B (PS B2) biosynthesis locus, complete sequence; and unknown genes.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
22689642_c2_249	1006	6228	355	1068	465	4.7e-44

Protein name	Locus Name	Acc#
putative glycosyltransferase	gp:AF125164	AF125164

Description

Bacteroides fragilis 638R polysaccharide B (PS B2) biosynthesis locus, complete sequence; and unknown genes.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
23490875_c3_289	1007	6229	508	1527	195	3.4e-12

Protein name	Locus Name	Acc#
putative flippase	gp:AF125164	AF125164

Description

Bacteroides fragilis 638R polysaccharide B (PS B2) biosynthesis locus, complete sequence; and unknown genes.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
23553136_c2_238	1008	6230	345	1038		

Protein name	Locus Name	Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
23554555_f3_142	1009	6231	254	765		

Protein name _____ Locus Name _____ Acc# _____

Description _____

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
23595137_f3_115	1010	6232	119	360		

Protein name _____ Locus Name _____ Acc# _____

Description _____

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
23635952_c1_182	1011	6233	333	1002	123	6.7e-11

Protein name _____ Locus Name _____ Acc# _____

dolichol-P-glucose synthetase homolog pir:E69322 E69322

Description _____

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
23678252_c2_247	1012	6234	444	1335	1376	1.4e-140

Protein name _____ Locus Name _____ Acc# _____

phosphoenolpyruvate phosphomutase FOM1 pir:S60206

Description _____

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
2382882_c2_216	1013	6235	383	1152	403	1.7e-37

Protein name _____ Locus Name _____ Acc# _____

hypothetical protein pir:S76344 S76344

Description _____

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24017687_f3_152	1014	6236	284	855	304	5.4e-27

Protein name

Locus Name

Acc#

sp:CDSA_HAEIN

Description

SYNTHASE)

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24226587_c2_241	1015	6237	314	945	301	1.1e-26

Protein name

Locus Name

Acc#

activator protein

gp:AF047527

AF047527

Description

Pseudomonas fluorescens activator protein (mtIR) gene, completecds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24398376_f3_148	1016	6238	107	324		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24413577_f1_44	1017	6239	288	867		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
-----------------	-------------	-------------	------------------	------------------	--------------	--------------------

24641925_c1_203	1018	6240	401	1206	136	3.8e-06
-----------------	------	------	-----	------	-----	---------

Protein name

Locus Name

Acc#

galactosyltransferase homolog

pir:G69465

G69465

Description

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
-----------------	-------------	-------------	------------------	------------------	--------------	--------------------

24806538_c3_292	1019	6241	342	1029	179	2.0e-11
-----------------	------	------	-----	------	-----	---------

Protein name

Locus Name

Acc#

capsular polysaccharide biosynthesis protein

pir:F70441

F70441

Description

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
-----------------	-------------	-------------	------------------	------------------	--------------	--------------------

2537575_c2_234	1020	6242	124	375	87	0.011
----------------	------	------	-----	-----	----	-------

Protein name

Locus Name

Acc#

probable membrane protein
YOL019w:hypothetical protein O2313

pir:S66701

S66701

Description

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
-----------------	-------------	-------------	------------------	------------------	--------------	--------------------

26204682_f3_128	1021	6243	326	981	221	3.3e-18
-----------------	------	------	-----	-----	-----	---------

Protein name

Locus Name

Acc#

sp:Y266_ARCFU

O29973

Description

HYPOTHETICAL PROTEIN AF0266

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
26367176_c2_217	1022	6244	251	756		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
26379677_f1_43	1023	6245	300	903	539	6.7e-52

Protein name

Locus Name

Acc#

sp:YGI2_PSEPU

P31857

Description

HYPOTHETICAL 32.4 KD PROTEIN IN GIDB-UNCI INTERGENIC REGION

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
29850282_f1_19	1024	6246	269	810	216	1.7e-17

Protein name

Locus Name

Acc#

sp:Y665_HAEIN

P44033

Description

HYPOTHETICAL PROTEIN HI0665

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
32636311_f2_83	1025	6247	480	1443		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
33376906_c3_290	1026	6248	316	951	351	5.6e-32
Protein name			Locus Name		Acc#	
LicD1			gp:AF106539		AF106539	
Description						
Streptococcus pneumoniae LicD1 (licD1) and LicD2 (licD2) genes, complete cds; and unknown gene.						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
33406567_f2_82	1027	6249	925	2778	129	6.1e-05
Protein name			Locus Name		Acc#	
115K outer membrane protein precursor:SusC protein			pir:JC6027		JC6027	
Description						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
33839461_c1_200.....	1028	6250	338	1017	348	1.2e-31
Protein name			Locus Name		Acc#	
putative alcohol dehydrogenase			gp:CZA382		AL078635	
Description						
Amycolatopsis orientalis cosmid pCZA382.						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
35396876_c3_295.....	1029	6251	418	1257	2046	1.4e-211
Protein name			Locus Name		Acc#	
putative epimerase			gp:AF125164		AF125164	
Description						
Bacteroides fragilis 638R polysaccharide B (PS B2) biosynthesis locus, complete sequence; and unknown genes.						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
35401627_c3_288	1030	6252	141	426	431	1.9e-40

Protein name

Locus Name

Acc#

WcgF

gp:AF125164

AF125164

Description

Bacteroides fragilis 638R polysaccharide B (PS B2) biosynthesis locus, complete sequence; and unknown genes.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
36362675_c1_207	1031	6253	197	594	317	2.2e-28

Protein name

Locus Name

Acc#

gp:AB008550

AB008550

Description

Pseudomonas aeruginosa phage phi CTX, complete genome sequence.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
3914025_c3_287	1032	6254	166	501	181	5.8e-14

Protein name

Locus Name

Acc#

unknown

gp:AF125164

AF125164

Description

Bacteroides fragilis 638R polysaccharide B (PS B2) biosynthesis locus, complete sequence; and unknown genes.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
3943753_c2_245	1033	6255	296	891	1278	3.3e-130

Protein name

Locus Name

Acc#

glucose-1-phosphate thymidyltransferase

gp:AF125164

AF125164

Description

Bacteroides fragilis 638R polysaccharide B (PS B2) biosynthesis locus, complete sequence; and unknown genes.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
3955062_c3_299	1034	6256	265	798	918	4.6e-92
<u>Protein name</u>			<u>Locus Name</u>		<u>Acc#</u>	
unknown			gp:AF125164		AF125164	
<u>Description</u>						
Bacteroides fragilis 638R polysaccharide B (PS B2) biosynthesis locus, complete sequence; and unknown genes.						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
3991300_c3_258	1035	6257	295	888	405	1.1e-37
<u>Protein name</u>			<u>Locus Name</u>		<u>Acc#</u>	
stationary phase survival protein Sure			pir:A70372		A70372	
<u>Description</u>						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4157762_c2_244	1036	6258	182	549	95	0.00012
<u>Protein name</u>			<u>Locus Name</u>		<u>Acc#</u>	
unknown			gp:AF048749		AF048749	
<u>Description</u>						
Bacteroides fragilis capsular polysaccharide biosynthesis operon, complete sequence.						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4175255_f3_151	1037	6259	680	2043	1389	4.3e-148
<u>Protein name</u>			<u>Locus Name</u>		<u>Acc#</u>	
FtsH2			gp:AB023310		AB023310	
<u>Description</u>						
Cyanidioschyzon merolae gene for FtsH2, complete cds.						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4304812_c2_246	1038	6260	140	423	532	3.7e-51

Protein name

Locus Name

Acc#

WcgG

gp:AF125164

AF125164

Description

Bacteroides fragilis 638R polysaccharide B (PS B2) biosynthesis locus, complete sequence; and unknown genes.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4803555_c3_297	1039	6261	198	597	996	2.5e-100

Protein name

Locus Name

Acc#

putative acetyltransferase

gp:AF125164

AF125164

Description

Bacteroides fragilis 638R polysaccharide B (PS B2) biosynthesis locus, complete sequence; and unknown genes.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4878277_c1_192	1040	6262	205	618	95	0.0062

Protein name

Locus Name

Acc#

gp:YP102KB

AL031866

Description

Yersinia pestis 102 kbases unstable region: from 1 to 119443.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4897128_c1_201	1041	6263	298	897	299	1.8e-26

Protein name

Locus Name

Acc#

N-acetylglucosaminyltransferase

gp:AB017355

AB017355

Description

Streptococcus agalactiae DNA, cps (capsular polysaccharide) genes, partial and complete cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4897256_c1_183	1042	6264	498	1497	1152	7.4e-117

Protein name

Locus Name

Acc#

X-His dipeptidase,:aminoacylhistidine
dipeptidase:aminopeptidase
D:beta-alanyl-histidine

pir:JU0300

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4962760_c2_233	1043	6265	1927	5784	167	6.8e-20

Protein name

Locus Name

Acc#

hypothetical protein

pir:E72310

E72310

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
5136411_c1_202	1044	6266	362	1089	178	5.4e-11

Protein name

Locus Name

Acc#

capsular polysaccharide biosynthesis homolog
yveQ

pir:F70036

F70036

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
5213541_c3_263	1045	6267	281	846	297	3.0e-26

Protein name

Locus Name

Acc#

hypothetical protein APE2014

pir:H72504

H72504

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

5275281_f1_45	1046	6268	440	1323	375	1.3e-38
---------------	------	------	-----	------	-----	---------

Protein name

Locus Name

Acc#

probable membrane-bound lytic murein
transglycosylase D (dniR)

pir:H71301

H71301

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

6037801_c3_276	1047	6269	379	1140		
----------------	------	------	-----	------	--	--

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

6648452_f2_87	1048	6270	272	819	396	9.6e-37
---------------	------	------	-----	-----	-----	---------

Protein name

Locus Name

Acc#

sp:KSGA_MYCCA

P43038

Description

DIMETHYLTRANSFERASE)

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

6832757_f3_149	1049	6271	471	1416	665	3.0e-65
----------------	------	------	-----	------	-----	---------

Protein name

Locus Name

Acc#

Ykok

gp:AB013374

AB013374

Description

Bacillus halodurans C-125 mamX, yjda, ykok and yvfK genes, partial and complete cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
6853387_c1_198	1050	6272	367	1104	1050	4.8e-106
<u>Protein name</u>			<u>Locus Name</u>			<u>Acc#</u>
PCZA361.5			gp:AOPCZA361			AJ223998
<u>Description</u>						
Amycolatopsis orientalis cosmid PCZA361.						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
800812_c2_235	1051	6273	379	1140		
<u>Protein name</u>			<u>Locus Name</u>			<u>Acc#</u>
<u>Description</u>						
NO-HIT						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
824051_c1_206	1052	6274	402	1209	1943	1.1e-200
<u>Protein name</u>			<u>Locus Name</u>			<u>Acc#</u>
putative aminotransferase			gp:AF125164			AF125164
<u>Description</u>						
Bacteroides fragilis 638R polysaccharide B (PS B2) biosynthesis locus, complete sequence; and unknown genes.						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
882702_c2_237	1053	6275	279	840	114	1.3e-06
<u>Protein name</u>			<u>Locus Name</u>			<u>Acc#</u>
unknown			gp:AF068902			AF068902
<u>Description</u>						
Streptococcus pneumoniae D-glutamic acid adding enzyme MurD (murD), undecaprenyl-PP-MurNAc-pentapeptide-UDPGlcNAc GlcNAc transferase (murG), cell division protein DivIB (divIB), orotidine-5'-decarboxylase PyrF (pyrF), and orotate phosphoribosyltransferase PyrE (pyrE) genes, complete cds; and unknown						

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
-----------------	-------------	-------------	------------------	------------------	--------------	--------------------

99444428_f3_97	1054	6276	100	303		
----------------	------	------	-----	-----	--	--

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
-----------------	-------------	-------------	------------------	------------------	--------------	--------------------

24864688_c1_7	1055	6277	77	234		
---------------	------	------	----	-----	--	--

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
-----------------	-------------	-------------	------------------	------------------	--------------	--------------------

29317557_c2_9	1056	6278	519	1560	142	6.3e-06
---------------	------	------	-----	------	-----	---------

Protein name

Locus Name

Acc#

receptor antigen (RagA)

gp:PGI130872

AJ130872

Description

Porphyromonas gingivalis W50 receptor antigen (rag) locus encoding a major immunodominant 55kDa antigen.

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
-----------------	-------------	-------------	------------------	------------------	--------------	--------------------

1038461_f3_25	1057	6279	274	825		
---------------	------	------	-----	-----	--	--

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
16211377_f3_28	1058	6280	158	477		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
16486287_f1_6	1059	6281	151	456		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
24647938_f1_5	1060	6282	133	402	110	2.2e-06

Protein name

Locus Name

Acc#

muramoyl-pentapeptide carboxypeptidase

pir:T34747

T34747

Description

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
24666313_f3_31	1061	6283	948	2847	218	6.3e-14

Protein name

Locus Name

Acc#

slow myosin heavy chain 2

gp:GGU85023

U85023

Description

Gallus gallus slow myosin heavy chain 2 (SM2) mRNA, partial cds.

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
4100885_f3_26	1062	6284	316	951		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
4119637_f1_8	1063	6285	215	648	115	0.00017

Protein name

Locus Name

Acc#

hypothetical protein jhp0052

pir:F71980

F71980

Description

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
4900252_f1_1	1064	6286	264	795		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
5946032_f2_16	1065	6287	285	858		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

10755437_f1_11	1066	6288	168	507	280	1.9e-24
----------------	------	------	-----	-----	-----	---------

Protein name

Locus Name

Acc#

sp:BKDR_PSEPU

P42179

Description

BKD OPERON TRANSCRIPTIONAL REGULATOR

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

1175211_f3_35	1067	6289	215	648	550	4.6e-53
---------------	------	------	-----	-----	-----	---------

Protein name

Locus Name

Acc#

inner membrane ABC transporter

gp:AF213822

AF213822

Description

Zymomonas mobilis strain ZM4 fosmid clone 42B3, complete sequence.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

12272127_f3_40	1068	6290	140	423		
----------------	------	------	-----	-----	--	--

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

1423427_f1_4	1069	6291	608	1827	871	4.4e-87
--------------	------	------	-----	------	-----	---------

Protein name

Locus Name

Acc#

gp:YP102KB

AL031866

Description

Yersinia pestis 102 kbases unstable region: from 1 to 119443.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

15657687_f1_13	1070	6292	185	558	373	2.6e-34
----------------	------	------	-----	-----	-----	---------

Protein name

Locus Name

Acc#

sp:YBDM_ECOLI

P77174

Description

HYPOTHETICAL 23.9 KD PROTEIN IN CSTA-DSBG INTERGENIC REGION

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

15900317_f1_5	1071	6293	337	1014	315	1.6e-27
---------------	------	------	-----	------	-----	---------

Protein name

Locus Name

Acc#

NrpB

gp:PMU46488

U46488

Description

Proteus mirabilis NrpS (nrpS) gene, partial cds, NrpU (nrpU), NrpT (nrpT), NrpA (nrpA), NrpB (nrpB), NrpG (nrpG) and IrpP (irpP) genes, complete cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

20980080_f1_15	1072	6294	117	354	135	4.0e-08
----------------	------	------	-----	-----	-----	---------

Protein name

Locus Name

Acc#

60kDa protein

gp:AB004560

AB004560

Description

Porphyromonas gingivalis DNA for 60kDa protein, complete cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

22657052_c1_54	1073	6295	156	471		
----------------	------	------	-----	-----	--	--

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
22928450_f1_16	1074	6296	60	183		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
23556577_f2_25	1075	6297	434	1305	537	1.2e-59

Protein name

Locus Name

Acc#

sp:YBDN_ECOLI

P77216

Description

HYPOTHETICAL 47.8 KD PROTEIN IN CSTA-DSBG INTERGENIC REGION

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
23867917_c2_88	1076	6298	96	291	84	0.0043

Protein name

Locus Name

Acc#

MHC class II alpha chain

gp:AF091557

AF091557

Description

Aulonocara hansbaenschi MHC class II alpha chain MHC-Auha-DAA1 mRNA (MHC-Auha-DAA1*01 allele), complete cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24431537_c2_75	1077	6299	108	327		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24492078_f1_1	1078	6300	307	924		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
29580387_f3_38	1079	6301	317	954	142	1.7e-07

Protein name

Locus Name

Acc#

pobR regulator

gp:PSEY18527

Y18527

Description

Pseudomonas sp. pobA, pobR, pcaQ, pcaH and pcaG genes.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
31691875_f2_19	1080	6302	214	645	166	4.7e-12

Protein name

Locus Name

Acc#

gp:LIINLC

Y07639

Description

L.ivanovii 23S rRNA, 5S rRNA, tRNA-Asn, tRNA-Thr, ORF Z, inlD, and inlC genes.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
33399057_c1_52	1081	6303	288	867	159	1.5e-09

Protein name

Locus Name

Acc#

sp:LCRF_YERPE

P28808

Description

THERMOREGULATORY PROTEIN LCRF

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

35650462_f3_43	1082	6304	200	603	272	4.8e-23
----------------	------	------	-----	-----	-----	---------

Protein name

Locus Name

Acc#

60kDa protein

gp:AB004560

AB004560

Description

Porphyromonas gingivalis DNA for 60kDa protein, complete cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

4069180_f1_14	1083	6305	197	594		
---------------	------	------	-----	-----	--	--

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

4953586_c2_89	1084	6306	119	360	92	0.0026
---------------	------	------	-----	-----	----	--------

Protein name

Locus Name

Acc#

lipase precursor

gp:AF053006

AF053006

Description

Staphylococcus epidermidis lipase precursor (geh1) gene, complete cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

4954462_f1_3	1085	6307	275	828	265	7.3e-23
--------------	------	------	-----	-----	-----	---------

Protein name

Locus Name

Acc#

sp:TCMP_STRGA

P39887

Description

(EC 2.1.1.-)

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
5260317_c2_80	1086	6308	61	186	54	0.042
<u>Protein name</u>	<u>Locus Name</u>				<u>Acc#</u>	
pqqG protein	pir:B55527				B55527	
<u>Description</u>						

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
7065802_l3_28	1087	6309	238	717	81	0.0088
<u>Protein name</u>	<u>Locus Name</u>				<u>Acc#</u>	
hypothetical protein MTH1102	pir:F69013				F69013	
<u>Description</u>						

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
9963202_c3_101	1088	6310	511	1536	373	1.2e-32
<u>Protein name</u>	<u>Locus Name</u>				<u>Acc#</u>	
sensory transduction histidine kinase sll0474:protein sll0474:protein sll0474	pir:S76650				S76650	
<u>Description</u>						

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
11723417_c2_40	1089	6311	484	1455	121	2.8e-06
<u>Protein name</u>	<u>Locus Name</u>				<u>Acc#</u>	
unknown	gp:U96771				U96771	
<u>Description</u>						
Prevotella bryantii putative polygalacturonase, B-1,4-endoglucanase, and mannanase genes, complete cds; and unknown genes.						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
16601526_c1_38	1090	6312	1089	3270	801	7.4e-91
<u>Protein name</u>			<u>Locus Name</u>			<u>Acc#</u>
receptor antigen (RagA)			gp:PGI130872			AJ130872
<u>Description</u>						
Porphyromonas gingivalis W50 receptor antigen (rag) locus encoding a major immunodominant 55kDa antigen.						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
20507937_c2_42	1091	6313	542	1629	138	3.6e-07
<u>Protein name</u>			<u>Locus Name</u>			<u>Acc#</u>
unknown			gp:U96771			U96771
<u>Description</u>						
Prevotella bryantii putative polygalacturonase, B-1,4-endoglucanase, and mannanase genes, complete cds; and unknown genes.						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
21678137_f3_27	1092	6314	420	1263	101	0.024
<u>Protein name</u>			<u>Locus Name</u>			<u>Acc#</u>
hypothetical protein ytaP			pir:B69988			B69988
<u>Description</u>						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24491512_c3_49	1093	6315	543	1632	147	5.0e-07
<u>Protein name</u>			<u>Locus Name</u>			<u>Acc#</u>
unknown			gp:U96771			U96771
<u>Description</u>						
Prevotella bryantii putative polygalacturonase, B-1,4-endoglucanase, and mannanase genes, complete cds; and unknown genes.						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
35351583_f3_34	1094	6316	71	216		

Protein name

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4120307_f3_26	1095	6317	443	1332	520	6.9e-50

Protein name

hypothetical protein PAB1371

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4144515_f2_17	1096	6318	70	213		

Protein name

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
781932_c1_35	1097	6319	1100	3303	543	2.7e-85

Protein name

115K outer membrane protein precursor:Susc protein

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
22697711_c3_9	1098	6320	329	990	614	9.2e-68
Protein name					Locus Name	Acc#
neuraminidase precursor					gp:BNRNANASE	D28493
Description						
Bacteroides fragilis nanH gene for neuraminidase, complete cds.						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
11182950_c2_139	1099	6321	248	747	619	2.2e-60
Protein name					Locus Name	Acc#
TruB					gp:AF169967	AF169967
Description						
Flavobacterium johnsoniae leuS (leuS) gene, partial cds; and Fjoi2(fjoi2), FtsX (ftsX), Fjoi3 (fjoi3), BacA (bacA), and TruB (truB) genes, complete cds.						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
11956503_c1_120.....	1100	6322	142	429	110	1.9e-06
Protein name					Locus Name	Acc#
					sp:RNPA_BORBU	P50069
Description						
RIBONUCLEASE P PROTEIN COMPONENT, (PROTEIN C5) (RNASE P)						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
13864002_f1_26.....	1101	6323	73	222		
Protein name					Locus Name	Acc#
Description						
NO-HIT						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
14882902_c1_108	1102	6324	381	1146	117	0.00018
<u>Protein name</u>			<u>Locus Name</u>		<u>Acc#</u>	
sensory transduction system regulatory protein slr1837:protein slr1837:protein slr1837			pir:S77341		S77341	
<u>Description</u>						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
15109377_c1_119	1103	6325	254	765	108	0.0011
<u>Protein name</u>			<u>Locus Name</u>		<u>Acc#</u>	
			sp:HEM4_SCHPO		P87214	
<u>Description</u>						
(URO111S)						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
20176878_c3_156.....	1104	6326	280	843	106	0.00063
<u>Protein name</u>			<u>Locus Name</u>		<u>Acc#</u>	
ATPase subunit 6			gp:TCU40265		U40265	
<u>Description</u>						
Trypanosoma cruzi ATPase subunit 6 mRNA, complete cds.						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
204775_c3_153.....	1105	6327	304	915	466	3.7e-44
<u>Protein name</u>			<u>Locus Name</u>		<u>Acc#</u>	
FtsX			gp:AF169967		AF169967	
<u>Description</u>						
Flavobacterium johnsoniae LeuS (leuS) gene, partial cds; and Fjo12(fjo12), FtsX (ftsX), Fjo13 (fjo13), BacA (bacA), and TruB (truB)genes, complete cds.						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
2116037_c1_107	1106	6328	76	231		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
21522150_c2_141	1107	6329	437	1314	597	1.4e-110

Protein name

Locus Name

Acc#

sp:METK_HAEIN

P43762

Description

ADENOSYLTRANSFERASE) (ADOMET SYNTHETASE)

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
23442175_c1_122	1108	6330	439	1320	1071	2.8e-108

Protein name

Locus Name

Acc#

sp:SYN_BACST

P00952

Description

TYROSYL-TRNA SYNTHETASE, (TYROSINE--TRNA LIGASE) (TYRRS)

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
23447031_c1_109	1109	6331	298	897		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
23453160_f1_10	1110	6332	60	183		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
23712752_c3_149.....	1111	6333	89	270		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
24256300_c2_138.....	1112	6334	676	2031		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
24406253_c3_163.....	1113	6335	93	279	142	1.5e-09

Protein name

Locus Name

Acc#

oxidoreductase, short chain
dehydrogenase/reductase family

pir:A72395

A72395

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

26605287_c1_114	1114	6336	268	807	400	3.6e-37
-----------------	------	------	-----	-----	-----	---------

Protein name

Locus Name

Acc#

sp:BACA_ECOLI

Description

(EC 2.7.1.66)

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

29303427_f1_14	1115	6337	81	246	70	0.033
----------------	------	------	----	-----	----	-------

Protein name

Locus Name

Acc#

hypothetical protein A635R

pir:T18137

T18137

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

29932918_c3_151.....	1116	6338	85	258		
----------------------	------	------	----	-----	--	--

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

32228430_c3_161.....	1117	6339	527	1584	447	3.8e-42
----------------------	------	------	-----	------	-----	---------

Protein name

Locus Name

Acc#

choline sulfatase

gp:RMU39940

U39940

Description

Sinorhizobium meliloti bet operon, complete sequence.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
34251637_f2_70	1118	6340	90	273		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
35791416_c1_105.....	1119	6341	692	2079	449	2.4e-39

Protein name

Locus Name

Acc#

putative secreted beta-galactosidase

gp:SCF81

AL133171

Description

Streptomyces coelicolor cosmid F81.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
36366552_c1_113.....	1120	6342	79	240	175	2.5e-13

Protein name

Locus Name

Acc#

Fjo13

gp:AF169967

AF169967

Description

Flavobacterium johnsoniae LeuS (leuS) gene, partial cds; and Fjo12(fjo12), FtsX (ftsX), Fjo13 (fjo13), BacA (bacA), and TruB (truB)genes, complete cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
36437893_c3_148.....	1121	6343	67	204		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
3937750_c1_110	1122	6344	354	1065		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
3942082_c2_140.....	1123	6345	357	1074	753	1.4e-74

Protein name

Locus Name

Acc#

S-adenosylmethionine tRNA ribosyltransferase

pir:A72360

A72360

Description

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
4102177_c3_162.....	1124	6346	283	852	735	1.1e-72

Protein name

Locus Name

Acc#

sp:KDUI_ERWCH

Q05529

Description

(5-KETO-4-DEOXYURONATE ISOMERASE) (DKI ISOMERASE)

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
4119005_c3_158.....	1125	6347	212	639	159	1.6e-11

Protein name

Locus Name

Acc#

HI0454

gp:AF174390

AF174390

Description

Haemophilus influenzae strain Rd KW20 HI0454 gene, partial cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4188438_c3_157	1126	6348	76	231	226	9.9e-19
Protein name			Locus Name			Acc#
conserved hypothetical protein			pir:G72251			G72251
Description						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4398382_c1_118.....	1127	6349	201	606	374	2.1e-34
Protein name			Locus Name			Acc#
conserved hypothetical protein yvdD			pir:D70033			D70033
Description						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4757_c3_152.....	1128	6350	375	1128		
Protein name			Locus Name			Acc#
Description						
NO-HIT						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4867142_f3_74.....	1129	6351	1098	3297	302	1.1e-38
Protein name			Locus Name			Acc#
hypothetical protein c0624			pir:S73091			S73091
Description						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4876438_f3_102.....	1130	6352	590	1773	162	1.7e-09
Protein name			Locus Name			Acc#
response regulator			gp:SPAJ6398			AJ006398
Description						
Streptococcus pneumoniae rr09 and hk09 genes; two component system09.						

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
4876515_c2_134	1131	6353	376	1131		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
4954376_c2_137	1132	6354	355	1068		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
5273377_c2_132	1133	6355	359	1080		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
553161_c2_133	1134	6356	81	246		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
553161_c2_135	1135	6357	83	252		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
5892138_c1_106	1136	6358	409	1230	133	4.7e-06
Protein name			Locus Name		Acc#	
hypothetical protein PH0283			pir:D71453		D71453	
Description						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
6829630_c2_136	1137	6359	376	1131		
Protein name			Locus Name		Acc#	
Description						

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
6835381_f3_93	1138	6360	65	198		
Protein name			Locus Name		Acc#	
Description						

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
6837828_c1_117	1139	6361	154	465	326	2.5e-29
Protein name			Locus Name		Acc#	
			sp:HPPK_PORGI		083019	
Description						

(HPPK) (6-HYDROXYMETHYL-7,8-DIHYDROPTERIN PYROPHOSPHOKINASE) (PPPK)

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
7315641_c1_111	1140	6362	311	936	140	2.7e-07

Protein name

Locus Name

Acc#

ubiquinone/menaquinone biosynthesis
methyltransferase-related protein

pir:F72262

F72262

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
10978590_f3_64	1141	6363	77	234		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
14257762_f3_66	1142	6364	110	333		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
16453180_f3_65	1143	6365	150	453	50	0.020

Protein name

Locus Name

Acc#

WW domain binding protein 5

gp:MMU92454

U92454

Description

Mus musculus WW domain binding protein 5 mRNA, partial cds.

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
2084768_f3_57	1144	6366	149	450		

<u>Protein name</u>	<u>Locus Name</u>	<u>Acc#</u>

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
23705002_f1_10	1145	6367	746	2241	243	1.2e-29

<u>Protein name</u>	<u>Locus Name</u>	<u>Acc#</u>

conserved hypothetical protein ylbK	pir:H69874	H69874
-------------------------------------	------------	--------

Description

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
24422676_f1_15	1146	6368	1023	3072		

<u>Protein name</u>	<u>Locus Name</u>	<u>Acc#</u>

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
24492125_c1_105	1147	6369	1025	3075	2141	1.2e-221

<u>Protein name</u>	<u>Locus Name</u>	<u>Acc#</u>

hypothetical protein mexF	pir:T30830	T30830
---------------------------	------------	--------

Description

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
24643800_f2_37	1148	6370	258	777	572	2.1e-55

<u>Protein name</u>	<u>Locus Name</u>	<u>Acc#</u>

sp:YAFV_ECOLI	Q47679
---------------	--------

Description

HYPOTHETICAL 28.9 KD PROTEIN IN DNAQ-GMHA INTERGENIC REGION

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
25975307_f1_27	1149	6371	112	339		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
26680340_f1_16	1150	6372	125	378		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
26756551_c1_104	1151	6373	409	1230	485	3.5e-46

Protein name

Locus Name

Acc#

sp:ACRE_ECOLI

P24180

Description

ACRIFLAVIN RESISTANCE PROTEIN E PRECURSOR (ENVC PROTEIN)

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
30583162_c2_108	1152	6374	981	2946	1508	2.3e-166

Protein name

Locus Name

Acc#

transcription-repair coupling factor

gp:AF023181

AF023181

Description

Listeria monocytogenes transcription-repair coupling factor (mfdL), low temperature requirement B protein (ltrB), and DivIC homolog (divL) genes, complete cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
31375817_f2_44	1153	6375	113	342	69	0.042

Protein name conserved hypothetical protein AF0188
Locus Name pir:D69273
Acc# D69273
Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
32147151_f1_12	1154	6376	392	1179	444	7.8e-42

Protein name
Locus Name sp:NAGA_VIBCH
Acc# 032445
Description DEACETYLASE)

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
32282637_f2_36	1155	6377	196	591	310	1.2e-27

Protein name hypothetical protein
Locus Name pir:G75263
Acc# G75263
Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
34181513_f1_26	1156	6378	460	1383	293	4.1e-42

Protein name dihydroorotase (pyrc) PAB1149
Locus Name pir:C75027
Acc# C75027
Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
34430317_f2_38	1157	6379	262	789	304	5.4e-27

Protein name	Locus Name	Acc#
protein-tyrosine phosphatase	gp:AB028630	AB028630

Description

Clostridium perfringens hyp27, bach, ptp, cpd genes for hypothetical protein, bacterial hemoglobin, protein-tyrosine phosphatase, 2', 3'-cyclic nucleotide 2'-phosphodiesterase, partial and complete cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4554753_f2_45	1158	6380	161	486	211	3.8e-17

Protein name	Locus Name	Acc#
	sp:YQGC_BACSU	P54486

Description

HYPOTHETICAL 17.3 KD PROTEIN IN CCCA-SODA INTERGENIC REGION

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4570341_f1_25	1159	6381	263	792	531	4.7e-51

Protein name	Locus Name	Acc#
putative glycosyl transferase.	gp:SC6D7	AL133213

Description

Streptomyces coelicolor cosmid 6D7.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
47200_f1_13	1160	6382	689	2070	439	7.4e-41

Protein name	Locus Name	Acc#
	sp:NAGB_BACSU	035000

Description

PHOSPHATE DEAMINASE) (GNPDA) (GLCN6P DEAMINASE)

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4876090_c1_82	1161	6383	204	615	122	0.00012

Protein name

Locus Name

Acc#

sp:MFD_BACSU

P37474

Description

TRANSCRIPTION-REPAIR COUPLING FACTOR (TRCF)

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4876300_c1_88	1162	6384	65	198		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
7157576_f3_67.....	1163	6385	642	1929	886	1.1e-88

Protein name

Locus Name

Acc#

conserved hypothetical protein

pir:C72391

C72391

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
835175_f1_28.....	1164	6386	226	678	237	1.9e-18

Protein name

Locus Name

Acc#

sp:METH_HUMAN

Description

(METHIONINE SYNTHASE, VITAMIN-B12 DEPENDENT) (MS)

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
10437958_c3_133	1165	6387	135	408		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
10647055_f2_17.....	1166	6388	263	792	124	5.7e-06

Protein name

Locus Name

Acc#

transcription regulator, crp family

pir:F72285

F72285

Description

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
1406526_c2_128.....	1167	6389	417	1254	876	1.3e-87

Protein name

Locus Name

Acc#

sp:PATB_BACSU

Q08432

Description

PUTATIVE AMINOTRANSFERASE B,

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
14494530_f1_6.....	1168	6390	862	2589	169	1.2e-08

Protein name

Locus Name

Acc#

outer membrane assembly protein (asmA) RP347

pir:E71691

E71691

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
14648577_c1_93	1169	6391	368	1107	174	1.1e-10

Protein name

Locus Name

Acc#

transmembrane sensor

gp:AF051691

AF051691

Description

Pseudomonas aeruginosa stress factor A (psfA), ECF sigma factor (fiuI), transmembrane sensor (fiuR), and hydroxamate-type ferrisiderophore receptor (fiuA) genes, complete cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
14875635_c3_139	1170	6392	313	942	602	1.4e-58

Protein name

Locus Name

Acc#

conserved hypothetical protein ytqA

pir:D69999

D69999

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
20088937_f1_11	1171	6393	292	879	717	9.2e-71

Protein name

Locus Name

Acc#

lipoic acid synthase

pir:A75480

A75480

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
22478311_c3_145	1172	6394	149	450		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
22708153_c3_132	1173	6395	364	1095	317	2.2e-28

Protein name

Locus Name

Acc#

GldB

gp:AF158372

AF158372

Description

Flavobacterium johnsoniae hypothetical protein gene, partial cds;GldB (gldB), GldC (gldC), and hypothetical protein genes, completecds; and hypothetical protein gene, partial cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
23620910_c2_111	1174	6396	85	258		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24337765_f3_68	1175	6397	904	2712	432	9.6e-68

Protein name

Locus Name

Acc#

115K outer membrane protein precursor:SusC protein

pir:JC6027

JC6027

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24431510_c3_137	1176	6398	243	732	108	0.00085

Protein name

Locus Name

Acc#

hypothetical protein yvqF

pir:G70045

G70045

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24695187_c2_110	1177	6399	274	825	998	1.5e-100

Protein name

Locus Name

Acc#

sp:NAGE_BORBU

030564

Description

PHOSPHATE DEAMINASE) (GNPDA) (GLCN6P DEAMINASE)

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
2814055_c2_99	1178	6400	320	963	304	5.4e-27

Protein name

Locus Name

Acc#

enoyl-acyl carrier protein reductase

pir:H75330

H75330

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
2869825_c2_103	1179	6401	157	474	143	6.2e-10

Protein name

Locus Name

Acc#

hypothetical protein APE2345

pir:F72462

F72462

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
30470261_c3_143	1180	6402	369	1110	562	2.5e-54

Protein name

Locus Name

Acc#

O-acetylhomoserine sulfhydrylase

pir:D72324

D72324

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
3256642_c2_97	1181	6403	66	201		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
36360812_F2_36	1182	6404	374	1125		

Protein name Locus Name Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
36406537_F1_5	1183	6405	673	2022	344	2.8e-40

Protein name Locus Name Acc#

gp:SC9745

Description

S.cerevisiae chromosome XIII cosmid 9745.

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
3944711_F3_52	1184	6406	207	624	399	4.6e-37

Protein name Locus Name Acc#

probable translation factor yci0

pir:F64874

F64874

Description

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
3946886_F3_53	1185	6407	144	435	114	7.3e-07

Protein name Locus Name Acc#

maturation protein ppm32

gp:AF166485

AF166485

Description

Glycine max maturation protein ppm32 (PM32) mRNA, complete cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4007687_f1_10	1186	6408	745	2238	1937	4.8e-200

Protein name

Locus Name

Acc#

DPP IV

gp:AB008194

AB008194

Description

Porphyromonas gingivalis gene for DPP IV, complete cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4119037_c3_138	1187	6409	281	846	134	9.7e-07

Protein name

Locus Name

Acc#

two-component response regulator
lytT-involved

pir:B69655

B69655

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4960937_c2_109	1188	6410	402	1209	524	2.6e-50

Protein name

Locus Name

Acc#

hypothetical protein b2710

pir:B65051

B65051

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
6772836_f3_47	1189	6411	292	879	376	1.3e-34

Protein name

Locus Name

Acc#

conserved hypothetical protein ykrA

pir:C69862

C69862

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
969812_c3_144	1190	6412	204	615	164	3.7e-12

Protein name

Locus Name

Acc#

RNA polymerase ECF-type sigma factor homolog yhdM

pir:C69826

C69826

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
975405_c2_100.....	1191	6413	245	738	339	1.0e-30

Protein name

Locus Name

Acc#

sam-dependent methytransferase

pir:C72086

C72086

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
10011662_c1_209.....	1192	6414	458	1374	922	1.7e-92

Protein name

Locus Name

Acc#

sp:PRIA_BACSU

Description

PRIMOSOMAL PROTEIN N' (REPLICATION FACTOR Y)

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
10350927_f3_138.....	1193	6415	510	1533	172	5.8e-16

Protein name

Locus Name

Acc#

hypothetical protein MJ0749

pir:E64393

E64393

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
11767812_f1_20	1194	6416	265	798	281	1.5e-24
<u>Protein name</u>			<u>Locus Name</u>		<u>Acc#</u>	
two-component response regulator lytT-involved			pir:B69655		B69655	
<u>Description</u>						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
1204052_f1_26	1195	6417	280	843	131	3.5e-08
<u>Protein name</u>			<u>Locus Name</u>		<u>Acc#</u>	
			sp:YGEK_ECOLI		Q46791	
<u>Description</u>						

HYPOTHETICAL TRANSCRIPTIONAL REGULATOR IN KDUI-LYSS INTERGENIC REGION

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
13729582_f1_24	1196	6418	195	588	128	2.4e-08
<u>Protein name</u>			<u>Locus Name</u>		<u>Acc#</u>	
hypothetical protein			pir:C72325		C72325	
<u>Description</u>						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
13773262_f1_41	1197	6419	87	264		
<u>Protein name</u>			<u>Locus Name</u>		<u>Acc#</u>	
<u>Description</u>						

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
14252182_f1_45	1198	6420	212	639	222	2.6e-18
<u>Protein name</u>			<u>Locus Name</u>		<u>Acc#</u>	
resolvase			pir:S38652		S38652	
<u>Description</u>						

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
14354025_f1_1	1199	6421	447	1344		
<u>Protein name</u>			<u>Locus Name</u>		<u>Acc#</u>	
<u>Description</u>						

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
14570301_f2_66	1200	6422	65	198		
<u>Protein name</u>			<u>Locus Name</u>		<u>Acc#</u>	
<u>Description</u>						

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
14928462_c3_289	1201	6423	63	192		
<u>Protein name</u>			<u>Locus Name</u>		<u>Acc#</u>	
<u>Description</u>						

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
16673406_f2_95	1202	6424	419	1260		
<u>Protein name</u>			<u>Locus Name</u>		<u>Acc#</u>	
<u>Description</u>						

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
16832885_c1_170	1203	6425	431	1296	1723	2.3e-177

Protein name	Locus Name	Acc#
hypothetical protein	pir:JQ1020	JQ1020
Description		

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
19614050_c1_162	1204	6426	165	498		

Protein name	Locus Name	Acc#
Description		

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
19735306_f2_97	1205	6427	65	198	63	0.045

Protein name	Locus Name	Acc#
	sp:SRD2_CAEEL	Q21767
Description		

SRD-2 PROTEIN

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
19773437_f3_126	1206	6428	64	195		

Protein name	Locus Name	Acc#
Description		

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
20328267_c1_164	1207	6429	66	201		

Protein name Locus Name Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
20390875_E3_151	1208	6430	192	579	193	3.1e-15

Protein name Locus Name Acc#

conserved hypothetical protein pir:E72312 E72312

Description

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
20437675_C2_241	1209	6431	235	708		

Protein name Locus Name Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
20703426_E2_94	1210	6432	67	204		

Protein name Locus Name Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
20723160_c2_218	1211	6433	240	723	138	3.3e-09

Protein name Locus Name Acc#

conserved hypothetical protein HP0713 pir:A64609 A64609

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
20976426_f3_114	1212	6434	91	276	124	3.9e-07
Protein name			Locus Name		Acc#	
asparaginase homolog yccC			pir:F69754		F69754	
Description						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
21647925_f1_7	1213	6435	459	1380	1231	3.1e-125
Protein name			Locus Name		Acc#	
			sp:DCUB_HAEIN		P44855	
Description						
ANAEROBIC C4-DICARBOXYLATE TRANSPORTER DCUB						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
21992175_f3_117	1214	6436	327	984	84	0.0070
Protein name			Locus Name		Acc#	
putative transmembrane efflux protein.			gp:SCF91		AL132973	
Description						
Streptomyces coelicolor cosmid F91.						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
22860128_c2_219	1215	6437	83	252	64	0.031
Protein name			Locus Name		Acc#	
			sp:SPRC_XENLA		P36378	
Description						
(OSTEONECTIN) (ON) (BASEMENT MEMBRANE PROTEIN BM-40)						

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
23617137_c3_268	1216	6438	224	675	267	4.5e-23

Protein name

Locus Name

Acc#

sp:YJV7_YEAST

P40893

Description

HYPOTHETICAL 22.0 KD PROTEIN IN HXT11-HXT8 INTERGENIC REGION

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
23631252_c2_239	1217	6439	229	690		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
23631550_c1_174	1218	6440	123	372	125	5.0e-08

Protein name

Locus Name

Acc#

hypothetical protein MJ1618

pir:A64502

A64502

Description

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
23632787_c1_193	1219	6441	314	945	110	3.6e-05

Protein name

Locus Name

Acc#

probable integrase/recombinase

pir:B71194

B71194

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24009637_c3_256	1220	6442	76	231	67	0.025

Protein name

Locus Name

Acc#

sp:VE2_HP338

Q80910

Description

REGULATORY PROTEIN E2

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24066056_f3_155	1221	6443	91	276	134	5.5e-09

Protein name

Locus Name

Acc#

sp:CEBA_BACAM

P23939

Description

BAMHI CONTROL ELEMENT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24236057_f2_73	1222	6444	60	183		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24353391_f1_40	1223	6445	153	462		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT</u> <u>Length</u>	<u>AA</u> <u>Length</u>	<u>Score</u>	<u>Probability</u>
24394017_f3_153	1224	6446	63	192		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT</u> <u>Length</u>	<u>AA</u> <u>Length</u>	<u>Score</u>	<u>Probability</u>
24485932_c3_279.....	1225	6447	848	2547	106	0.0070

Protein name

Locus Name

Acc#

arylesterase

gp:AF044683

AF044683

Description

Agrobacterium radiobacter putative dihydrolipoamideS-acetyltransferase (dla) gene, partial cds; arylesterase (ada)gene, complete cds; and putative dihydrolipoamide dehydrogenase(dlh) gene, partial cds.

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT</u> <u>Length</u>	<u>AA</u> <u>Length</u>	<u>Score</u>	<u>Probability</u>
2461567_c3_281.....	1226	6448	73	222		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT</u> <u>Length</u>	<u>AA</u> <u>Length</u>	<u>Score</u>	<u>Probability</u>
24640807_f2_53.....	1227	6449	397	1194		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24882203_f1_29	1228	6450	62	189		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24882932_c1_175	1229	6451	178	537	348	1.2e-31

Protein name

Locus Name

Acc#

adaptive response regulatory protein

gp:AF047839

AF047839

Description

Pseudoalteromonas sp. S9 putative glucosyl hydrolase precursor and adaptive response regulatory protein (ada) genes, complete cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
2531640_c2_242	1230	6452	200	603	352	4.4e-32

Protein name

Locus Name

Acc#

unknown

gp:AF006034

AF006034

Description

Clostridium pasteurianum 1,3-propanediol dehydrogenase (dhaT) gene, complete cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
25578438_f2_86	1231	6453	78	237		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
25914666_f3_147	1232	6454	92	279	86	0.010
<u>Protein name</u>	<u>Locus Name</u>				<u>Acc#</u>	
probable serine-threonine-protein kinase	pir:T41341				T41341	
<u>Description</u>						

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
26178177_c2_221	1233	6455	172	519	126	3.9e-08
<u>Protein name</u>	<u>Locus Name</u>				<u>Acc#</u>	
hypothetical protein MTH847	pir:A69213				A69213	
<u>Description</u>						

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
26306637_c3_297	1234	6456	448	1347	303	3.0e-24
<u>Protein name</u>	<u>Locus Name</u>				<u>Acc#</u>	
	sp:PRIA_BACSU					
<u>Description</u>						
PRIMOSOMAL PROTEIN N' (REPLICATION FACTOR Y)						

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
26360312_c3_290	1235	6457	328	987		
<u>Protein name</u>	<u>Locus Name</u>				<u>Acc#</u>	
<u>Description</u>						
NO-HIT						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
26369028_f1_8	1236	6458	323	972	912	2.0e-91
Protein name			Locus Name		Acc#	
			sp:ASG2_ECOLI		P00805	
Description						
AMIDOHYDROLASE II) (L-ASNASE II) (COLASPASE)						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
26460937_f2_102	1237	6459	392	1179	590	2.6e-57
Protein name			Locus Name		Acc#	
mannose-1-phosphate guanylyltransferase			pir:H72303		H72303	
Description						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
26593800_c3_267.....	1238	6460	197	594	580	3.0e-56
Protein name			Locus Name		Acc#	
			sp:YJV8_YEAST		P40892	
Description						
(EC 2.3.1.-)						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
26595192_f1_22.....	1239	6461	462	1389	534	2.3e-51
Protein name			Locus Name		Acc#	
oxidoreductase, aldo/keto reductase family			pir:E72284		E72284	
Description						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
26601062_c3_259	1240	6462	61	186		

Protein name Locus Name Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
29328501_f3_131	1241	6463	66	201		

Protein name Locus Name Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
31723562_c2_248	1242	6464	385	1158		

Protein name Locus Name Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
3320802_c3_294	1243	6465	77	234		

Protein name Locus Name Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
33245255_c1_188	1244	6466	261	786	103	0.014

Protein name Locus Name Acc#

hypothetical protein 2

pir:S49113

S49113

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
34078300_f3_152	1245	6467	284	855	182	6.3e-12

Protein name	Locus Name	Acc#
integrase	gp:BFU75371	U75371

Description

Bacteroides fragilis transposon Tn4555 TnpA (tnpA), integrase(int), TnpC (tnpC), excisionase (xis), mobilization protein (mobA), and beta-lactamase (cfxA) genes, complete cds; and unknown genes.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
34104127_f3_141	1246	6468	808	2427	274	3.0e-20

Protein name	Locus Name	Acc#
	sp:IRGA_VIBCH	P27772

Description

IRON-REGULATED OUTER MEMBRANE VIRULENCE PROTEIN PRECURSOR

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
34260911_f3_150.....	1247	6469	166	501		

Protein name	Locus Name	Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
35314080_f2_65.....	1248	6470	716	2151		

Protein name	Locus Name	Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
35704786_f2_92	1249	6471	92	279	326	2.5e-29

Protein name	Locus Name	Acc#
integrase IntN1	gp:BUU51917	U51917

Description

Bacteroides uniformis insertion element NBU1 fragment, integraseIntN1 gene, complete cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
3938817_c2_247	1250	6472	480	1443	1319	1.5e-134

Protein name	Locus Name	Acc#
aspartate ammonia-lyase	gp:WSAJ2933	AJ002933

Description

Wolinella succinogenes aspA, dcuA genes and partial ansA gene.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
3940943_f2_74	1251	6473	355	1068	180	2.8e-11

Protein name	Locus Name	Acc#
AlgZ	gp:PAU52431	U52431

Description

Pseudomonas aeruginosa AlgR-cognate sensor AlgZ (algZ) gene, complete cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4000953_f2_87	1252	6474	301	906	147	3.1e-08

Protein name	Locus Name	Acc#
transcription regulator	gp:AF008220	AF008220

Description

Bacillus subtilis rrnB-dnaB genomic region.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4072187_c1_172	1253	6475	587	1764	483	9.3e-95

Protein name	Locus Name	Acc#
	sp:DXS_BACSU	P54523

Description

PROBABLE 1-DEOXYXYLULOSE-5-PHOSPHATE SYNTHASE (DXP SYNTHASE)

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
423162_c1_207	1254	6476	221	666		

Protein name	Locus Name	Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4345337_f3_110	1255	6477	401	1206	143	3.8e-09

Protein name	Locus Name	Acc#
	gp:ECASPA	X02307

Description

E. coli aspA gene for aspartase (L-aspartate ammonia-lyase) (EC4.3.1.1).

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4454000_c2_238	1256	6478	806	2421	171	1.0e-08

Protein name	Locus Name	Acc#
R27-2 protein	pir:T30296	T30296

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4960312_f1_46	1257	6479	523	1572	227	7.2e-16

Protein name	Locus Name	Acc#
putative integrase	gp:BA1242593	AJ242593

Description

Bacteriophage A118 complete genome.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
579712_f1_42	1258	6480	339	1020		

Protein name	Locus Name	Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
6728452...c3...293.....	1259	6481	87	264		

Protein name	Locus Name	Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
867688...f3...136.....	1260	6482	206	621	243	1.6e-20

Protein name	Locus Name	Acc#
hypothetical protein slr2078	pir:S77566	S77566

Description

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
9800466_c1_192	1261	6483	63	192		

<u>Protein name</u>	<u>Locus Name</u>	<u>Acc#</u>

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
994002_f2_99.....	1262	6484	197	594	92	0.00034

<u>Protein name</u>	<u>Locus Name</u>	<u>Acc#</u>

probable prefoldin subunit APE1440	pir:G72622	G72622
------------------------------------	------------	--------

Description

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
10647550_f2_40.....	1263	6485	812	2439	498	2.7e-45

<u>Protein name</u>	<u>Locus Name</u>	<u>Acc#</u>

putative transmembrane protein Wzc	gp:AF104912	AF104912
------------------------------------	-------------	----------

Description

Escherichia coli K30 capsule biosynthesis cluster, partial sequence.

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
10985663_f1_25.....	1264	6486	101	306	75	0.034

<u>Protein name</u>	<u>Locus Name</u>	<u>Acc#</u>

nuclear factor kappa-B2	gp:HSU20816	U20816
-------------------------	-------------	--------

Description

Human nuclear factor kappa-B2 (NF-KB2) gene, partial cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
133301_f2_28	1265	6487	588	1767	1406	9.0e-144

Protein name

Locus Name

Acc#

sp:SYQ_ECOLI

Description

(GLNRS)

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
1350033_f1_11	1266	6488	224	675		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
1381287_c2_91	1267	6489	172	519	470	1.4e-44

Protein name

Locus Name

Acc#

sp:TPX_MYCTU

P95282

Description

PROBABLE THIOL PEROXIDASE,

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
13859702_f3_54	1268	6490	379	1140	181	3.8e-11

Protein name

Locus Name

Acc#

transposase

sp:AF038866

AF038866

Description

Bacteroides fragilis transposon Tn5520 transposase (bipH) and mobilization protein BmpH (bmpH) genes, complete cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
13945437_c3_98	1269	6491	155	468	86	0.00077

Protein name

Locus Name

Acc#

sp:DBH_THEMA

P36206

Description

DNA-BINDING PROTEIN HU

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
15738937_f1_24	1270	6492	104	315		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
15909682_f3_52	1271	6493	212	639	137	1.9e-08

Protein name

Locus Name

Acc#

hypothetical protein Rv1624c

pir:F70558

F70558

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
23475200_f1_8	1272	6494	482	1449	179	1.1e-10

Protein name

Locus Name

Acc#

conserved hypothetical protein MTH72

pir:B69196

B69196

Description

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
24070786_c2_97	1273	6495	254	762	356	1.7e-32

Protein name

Locus Name

Acc#

sp:YQGH_BACSU

P46339

Description

REGION (ORF72)

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
26438887_c3_100	1274	6496	424	1275		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
29947188_f1_15	1275	6497	62	189		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
30563966_f1_20	1276	6498	153	462	121	1.3e-07

Protein name

Locus Name

Acc#

hypothetical protein

pir:T28682

T28682

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
34179077_f1_16	1277	6499	262	789	170	1.2e-10

Protein name

Locus Name

Acc#

sp:EPSA_BURSO

Q45407

Description

EPS I POLYSACCHARIDE EXPORT OUTER MEMBRANE PROTEIN EPSA PRECURSOR

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
36131937_c1_78	1278	6500	149	450	123	2.5e-07

Protein name

Locus Name

Acc#

phosphate-binding protein PstS

pir:H69097

H69097

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4176337_f3_55	1279	6501	470	1413	656	2.7e-64

Protein name

Locus Name

Acc#

GumD protein

pir:S67820

S67820

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4744002_c2_89	1280	6502	198	597	249	3.6e-21

Protein name

Locus Name

Acc#

hypothetical protein (repA 5' region)

pir:S30120

S30120

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4867127_f2_31	1281	6503	216	651	308	2.0e-27

Protein name

Locus Name

Acc#

DedA family protein

pir:B75253

B75253

Description

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
6721890_c3_99	1282	6504	163	492	270	2.1e-23
<u>Protein name</u>	<u>Locus Name</u>				<u>Acc#</u>	
N-acetylmuramoyl-L-alanine amidase homolog	pir:G64126				G64126	
<u>Description</u>						

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
7313162_f1_4	1283	6505	285	858	410	3.1e-38
<u>Protein name</u>	<u>Locus Name</u>				<u>Acc#</u>	
phosphate-binding protein PstS	pir:H69097				H69097	
<u>Description</u>						

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
1212751_c1_87	1284	6506	147	444		
<u>Protein name</u>	<u>Locus Name</u>				<u>Acc#</u>	
<u>Description</u>						
NO-HIT						

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
14103176_c3_114	1285	6507	681	2046	246	2.5e-34
<u>Protein name</u>	<u>Locus Name</u>				<u>Acc#</u>	
	gp:PGU60208				U60208	
<u>Description</u>						
Porphyromonas gingivalis orf1, orf2 and orf3 genes, complete cds.						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
179762_c2_94	1286	6508	399	1200	465	9.2e-62

Protein name

Locus Name

Acc#

sp:YBDG_ECOLI

Description

HYPOTHETICAL 46.6 KD PROTEIN IN PHEP-NFNB INTERGENIC REGION

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
22063387_c2_92	1287	6509	620	1863	378	9.5e-67

Protein name

Locus Name

Acc#

alpha-1,3/4-fucosidase precursor

gp:SSU39394

U39394

Description

Streptomyces sp. alpha-1,3/4-fucosidase precursor gene, completecds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
238255_c3_113	1288	6510	322	969	106	0.0027

Protein name

Locus Name

Acc#

sp:YEHT_ECOLI

Description

HYPOTHETICAL 27.9 KD PROTEIN IN MOLR-BGLX INTERGENIC REGION

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24397305_c2_98	1289	6511	91	276		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24406550_c2_99	1290	6512	209	630	106	1.0e-05

Protein name

Locus Name

Acc#

gp:GGU25741

U25741

Description

Group G streptococcus strain g6 emmL gene, partial cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
25428436_c2_105	1291	6513	288	864		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
25897507_f1_15	1292	6514	356	1071	105	0.038

Protein name

Locus Name

Acc#

probable extracellular nuclease

pir:D75625

D75625

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
26210912_c3_117	1293	6515	408	1227	109	0.0058

Protein name

Locus Name

Acc#

silent surface layer protein

gp:AF079365

AF079365

Description

Lactobacillus crispatus silent surface layer protein (cbsB) gene, partial cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
265878_f3_62	1294	6516	114	345	80	0.043

Protein name	Locus Name	Acc#
MAR binding filament-like protein 1:MFP1 protein	pir:T07111	T07111

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
2845427_c1_89	1295	6517	939	2820	132	0.00012

Protein name	Locus Name	Acc#
	sp:PFEA_PSEAE	Q05098

Description

FERRIC ENTEROBACTIN RECEPTOR PRECURSOR

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
30192086_c1_76	1296	6518	566	1701	831	7.7e-83

Protein name	Locus Name	Acc#
	sp:BGAL_THEMA	

Description

BETA-GALACTOSIDASE, (LACTASE)

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
31256287_c3_116	1297	6519	641	1926	730	3.9e-72

Protein name	Locus Name	Acc#
	pir:S13786	

DNA-directed DNA polymerase, III chain
dnaX:DNA polymerase III (gamma and tau
subunits) dnaX

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
34570437_f2_47	1298	6520	492	1479	1142	8.5e-116
Protein name			Locus Name		Acc#	
			sp:PEPD_ECOLI		P15288	
Description						
(PEPTIDASE D)						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
35990807_c1_79	1299	6521	222	669	611	1.6e-59
Protein name			Locus Name		Acc#	
transaldolase-related protein			pir:G72394		G72394	
Description						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4101555_c2_102	1300	6522	332	999	144	2.0e-09
Protein name			Locus Name		Acc#	
			gp:APU72238		U72238	
Description						
Anabaena PCC7120 ORFR1, ORFR2, ORFR3, ORFR4, and ORFR5 genes, complete sequences.						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
447825_c3_106	1301	6523	502	1509	977	2.6e-98
Protein name			Locus Name		Acc#	
			sp:BGAL_BACME		O52847	
Description						
BETA-GALACTOSIDASE, (LACTASE)						

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
4961691_f2_39	1302	6524	117	354		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
15112533_f3_19	1303	6525	355	1068	210	5.1e-15

Protein name

Locus Name

Acc#

probable proteinase PAB1960

pir:A75179

A75179

Description

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
15820341_f2_14	1304	6526	157	474		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
204811_f2_16	1305	6527	274	825	105	0.012

Protein name

Locus Name

Acc#

gp:ATAC012563

AC012563

Description

Arabidopsis thaliana chromosome I BAC T23K23 genomic sequence, complete sequence.

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
21664062_f1_5	1306	6528	168	507		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
26604687_f2_12	1307	6529	205	618		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
33456962_f1_2	1308	6530	293	882		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
4098437_f1_1	1309	6531	268	807	115	0.00062

Protein name

Locus Name

Acc#

sp:Y066_METJA

Q60377

Description

HYPOTHETICAL PROTEIN MJ0066

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
6837782_f1_4	1310	6532	62	189		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
9939142_f1_3	1311	6533	115	348		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
10041563_c1_120	1312	6534	63	192		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
10726552_c3_185	1313	6535	80	243		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
12500086_c1_119	1314	6536	64	195		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
13860653_c3_194	1315	6537	488	1467	1040	5.5e-105

Protein name

Locus Name

Acc#

cell division protein

gp:PAL249201

AJ249201

Description

Prevotella albensis ftsQ (partial), ftsA and ftsZ genes and ORF-fts(partial).

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
14298312_c1_131	1316	6538	489	1470	1275	6.8e-130

Protein name

Locus Name

Acc#

sp:MURC_PORGI

Q51831

Description

ACETYLMURANOYL-L-ALANINE SYNTHETASE)

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
14644152_c2_164.....	1317	6539	254	765	341	6.4e-31

Protein name

Locus Name

Acc#

FtsQ

gp:AB004555

AB004555

Description

Porphyromonas gingivalis genes for FtsQ, FtsA, FtsZ, complete cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
164651_c2_169.....	1318	6540	669	2010	3334	0.0

Protein name

Locus Name

Acc#

DNA gyrase B subunit

gp:AB017713

AB017713

Description

Bacteroides fragilis gyrB gene for DNA gyrase B subunit, completecds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

16593937_c1_127	1319	6541	435	1308	388	6.7e-36
-----------------	------	------	-----	------	-----	---------

Protein name

Locus Name

Acc#

sp:YLAO_BACSU

007639

Description

HYPOTHETICAL 43.7 KD PROTEIN IN NPPE-PYCA INTERGENIC REGION

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

16808437_c3_192	1320	6542	135	408	223	6.9e-18
-----------------	------	------	-----	-----	-----	---------

Protein name

Locus Name

Acc#

UDP-N-acetylmuramoylalanine-D-glutamate
ligase

pir:H70477

H70477

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

177253_c1_137	1321	6543	66	201	62	0.047
---------------	------	------	----	-----	----	-------

Protein name

Locus Name

Acc#

Orfs10c

gp:SCU42227

U42227

Description

Saccharomyces cerevisiae replicative mitochondrial DNA polymerase catalytic subunit (MIP1) gene, nuclear gene encoding mitochondrial protein, partial cds, and putative 10-formyl-tetrahydrofolate binding protein (FTB1) gene, complete cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

19808211_c1_135	1322	6544	181	546	289	2.1e-25
-----------------	------	------	-----	-----	-----	---------

Protein name

Locus Name

Acc#

hypothetical protein 1

pir:S70830

S70830

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
20010316_f2_72	1323	6545	67	204		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
2149013_c3_187	1324	6546	343	1032	614	7.6e-60

Protein name

Locus Name

Acc#

unknown

gp:EFU94707

U94707

Description

Enterococcus faecalis strain A24836 cell wall/cell division genecluster, yllB, yllC, yllD, pbpC, mraY, murD, murG, divlB, ftsA andftsZ genes, complete cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
2151417_f3_78	1325	6547	180	543	186	2.5e-13

Protein name

Locus Name

Acc#

sp:YGY4_HALSQ

P21562

Description

HYPOTHETICAL 80.2 KD PROTEIN IN THE 5' REGION OF GYRA AND GYRB (ORF 4)

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
22303400_c2_166	1326	6548	442	1329	1291	1.4e-131

Protein name

Locus Name

Acc#

cell division protein

gp:PAL249201

AJ249201

Description

Prevotella albensis ftsQ (partial), ftsA and ftsZ genes and ORF-fts(partial).

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
23468837_f3_100	1327	6549	171	516		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
23646942_c1_126.....	1328	6550	328	987	368	8.9e-34

Protein name

Locus Name

Acc#

sp:MURD_BACSU

Description

ADDING ENZYME)

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
24072177_c3_175.....	1329	6551	452	1359	730	3.9e-72

Protein name

Locus Name

Acc#

hypothetical protein

pir:S76527

S76527

Description

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
24413875_c2_167.....	1330	6552	61	186		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
24414077_f1_33	1331	6553	208	627		

Protein name Locus Name Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
245643_f3_75	1332	6554	140	423	146	4.8e-09

Protein name Locus Name Acc#

conserved hypothetical protein pir:H75460 H75460

Description

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
24658577_f2_68	1333	6555	113	342		

Protein name Locus Name Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
25832161_f2_41	1334	6556	880	2643		

Protein name Locus Name Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
2620187_f3_96	1335	6557	217	654		

Protein name Locus Name Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
26750336_f1_37	1336	6558	347	1044	1218	7.5e-124
Protein name			Locus Name			Acc#
hemolysin A			gp:PMU27587			U27587
Description						
Prevotella melaninogenica hemolysin A (phyA) gene, complete cds.						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
2931518_c3_189	1337	6559	486	1461	892	2.6e-89
Protein name			Locus Name			Acc#
UDP-MurNac-tripeptide synthetase			pir:E70450			E70450
Description						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
30166437_c3_191.....	1338	6560	82	249	75	1.5e-06
Protein name			Locus Name			Acc#
phospho-n-acetylmuramoyl-pentapeptide-transferase (mraY1) RP595			pir:E71664			E71664
Description						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
31534452_f1_40.....	1339	6561	595	1788	342	1.6e-28
Protein name			Locus Name			Acc#
conserved hypothetical protein aq_854			pir:B70374			B70374
Description						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
3166057_c2_163	1340	6562	389	1170	641	1.0e-62

Protein name

Locus Name

Acc#

sp:MURG_BACSU

Description

(EC 2.4.1.-)

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
33398557_f3_102	1341	6563	206	621		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
33787927_f1_39.....	1342	6564	150	453	419	3.5e-39

Protein name

Locus Name

Acc#

sp:DUT_AQUAE

066592

Description

(DUTPASE) (DUTP PYROPHOSPHATASE)

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
33986038_f1_35.....	1343	6565	723	2172	164	9.8e-15

Protein name

Locus Name

Acc#

putative TonB-dependent outer membrane receptor

gp:AF048749

AF048749

Description

Bacteroides fragilis capsular polysaccharide biosynthesis operon, complete sequence.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

34260912_c1_122	1344	6566	118	357	75	0.0099
-----------------	------	------	-----	-----	----	--------

Protein name

Locus Name

Acc#

hypothetical protein 2

pir:I40759

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

37542_c2_153	1345	6567	198	597	229	4.8e-19
--------------	------	------	-----	-----	-----	---------

Protein name

Locus Name

Acc#

probable RNA polymerase sigma factor

pir:T42015

T42015

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

390927_f1_11	1346	6568	168	507	190	2.7e-14
--------------	------	------	-----	-----	-----	---------

Protein name

Locus Name

Acc#

gp:AB028868

AB028868

Description

Mus musculus P4(21)n mRNA, partial cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

3992135_f3_95	1347	6569	122	369		
---------------	------	------	-----	-----	--	--

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

4079668_f2_66	1348	6570	414	1245	105	0.032
---------------	------	------	-----	------	-----	-------

Protein name

Locus Name

Acc#

RING finger protein

gp:AF036255

AF036255

Description

Rattus norvegicus RING finger protein mRNA, complete cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

4174013_f1_7	1349	6571	265	798	101	6.7e-05
--------------	------	------	-----	-----	-----	---------

Protein name

Locus Name

Acc#

RecO

gp:HIU17037

U17037

Description

Haemophilus influenzae opacity associated proteins OapA and OapB(oapA and oapB) genes, complete cds, and DNA recombination andrepair protein (recO) gene, partial cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

48.758.12_c3_186.....	1350	6572	171	516	84	0.0019
-----------------------	------	------	-----	-----	----	--------

Protein name

Locus Name

Acc#

DNA-binding protein HB:DNA-binding protein HU:DNA-binding protein II

pir:S00015

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

49.578.37_f2_65.....	1351	6573	352	1059		
----------------------	------	------	-----	------	--	--

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
5117268_c1_121	1352	6574	160	483	177	1.5e-13

Protein name

Locus Name

Acc#

sp:YABB_ECOLI

P22186

Description

HYPOTHETICAL 17.4 KD PROTEIN IN FRUR-FTSL INTERGENIC REGION (ORFC)

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
5994087_c1_123	1353	6575	708	2127	337	2.1e-34

Protein name

Locus Name

Acc#

sp:SP5D_BACSU

Q03524

Description

BINDING PROTEIN)

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
6072683_c1_125	1354	6576	376	1131	344	8.1e-51

Protein name

Locus Name

Acc#

sp:MRAY_BORBU

Q44776

Description

(UDP-MURNAC-PENTAPEPTIDE PHOSPHOTRANSFERASE)

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
6649037_c2_168	1355	6577	89	270	153	5.4e-11

Protein name

Locus Name

Acc#

probable ribosomal protein S20 rpS1

pir:G70684

G70684

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
6758437_f1_36	1356	6578	286	861	186	7.0e-13

Protein name

Locus Name

Acc#

probable sulfolipid biosynthesis protein SqdA

pir:A42380

A42380

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
10328140_f1_3	1357	6579	71	216		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
10442793_f3_46	1358	6580	115	348		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
10744192_c1_95	1359	6581	281	846	153	8.1e-08

Protein name

Locus Name

Acc#

potassium channel alpha subunit Kv2.2

gp:XLU20342

U20342

Description

Xenopus laevis potassium channel alpha subunit Kv2.2 (XShab12)mRNA, complete cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

12518961_c2_100	1360	6582	325	978	241	5.2e-28
-----------------	------	------	-----	-----	-----	---------

Protein name

Locus Name

Acc#

probable protoporphyrinogen oxidase (hemK)
RP847

pir:G71646

G71646

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

14256430_f2_37	1361	6583	191	576	445	6.1e-42
----------------	------	------	-----	-----	-----	---------

Protein name

Locus Name

Acc#

conserved hypothetical protein MTH700

pir:E69193

E69193

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

14508510_f3_76	1362	6584	62	186		
----------------	------	------	----	-----	--	--

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

15835436_c3_152	1363	6585	308	927	228	6.1e-19
-----------------	------	------	-----	-----	-----	---------

Protein name

Locus Name

Acc#

hypothetical protein yitL

pir:E69840

E69840

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

16892517_c3_134	1364	6586	70	213		
-----------------	------	------	----	-----	--	--

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

19953510_c2_102	1365	6587	457	1374	626	4.1e-61
-----------------	------	------	-----	------	-----	---------

Protein name

Locus Name

Acc#

argininosuccinate lyase

pir:D70419

D70419

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

23673455_c1_80	1366	6588	161	486	126	3.9e-08
----------------	------	------	-----	-----	-----	---------

Protein name

Locus Name

Acc#

sp:RECX_PSEAE

P37860

Description

REGULATORY PROTEIN RECX

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

24026562_f3_59	1367	6589	415	1248	567	7.2e-55
----------------	------	------	-----	------	-----	---------

Protein name

Locus Name

Acc#

sp:ASSY_METJA

Q60174

Description

LIGASE)

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

24300018_c3_133	1368	6590	166	501		
-----------------	------	------	-----	-----	--	--

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

24353376_f1_6	1369	6591	533	1602	1279	2.6e-130
---------------	------	------	-----	------	------	----------

Protein name

Locus Name

Acc#

gp:AB024946

AB024946

Description

Escherichia coli plasmid pB171 DNA, complete sequence.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

25663952_f1_2	1370	6592	188	567	280	1.9e-24
---------------	------	------	-----	-----	-----	---------

Protein name

Locus Name

Acc#

sp:MTGA_ACICA

024849

Description

(EC 2.4.2.-) (MONOFUNCTIONAL TGASE)

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

25975187...c3...153.....	1371	6593	221	666		
--------------------------	------	------	-----	-----	--	--

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

26369087...f2...44.....	1372	6594	347	1044	554	1.7e-53
-------------------------	------	------	-----	------	-----	---------

Protein name

Locus Name

Acc#

riboflavin-specific deaminase

pir:G72207

G72207

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
32228388_c2_122	1373	6595	252	759	113	0.0016

Protein name

Locus Name

Acc#

sp:HEXA_BLADI

Q17127

Description

HEXAMERIN PRECURSOR

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
33367175_c3_131	1374	6596	163	492		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
36211443_f2_35	1375	6597	326	981	548	7.5e-53

Protein name

Locus Name

Acc#

N-acetyl-gamma-glutamyl-phosphate reductase,

pir:F69508

F69508

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4007801_c2_121	1376	6598	228	687		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

4095050_f3_61	1377	6599	260	783	375	1.6e-34
---------------	------	------	-----	-----	-----	---------

Protein name

Locus Name

Acc#

pyrroline-5-carboxylate reductase

gp:CSAJ10739

AJ010739

Description

Clostridium sticklandii proC gene and 5' flanking region.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

4377005_c3_130	1378	6600	230	693	541	4.1e-52
----------------	------	------	-----	-----	-----	---------

Protein name

Locus Name

Acc#

sp:PYRE_BACSU

P25972

Description

OROTATE PHOSPHORIBOSYLTRANSFERASE, (OPRT) (OPRTASE)

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

4801552_f2_45.....	1379	6601	458	1377		
--------------------	------	------	-----	------	--	--

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

4804051_f2_34.....	1380	6602	203	612		
--------------------	------	------	-----	-----	--	--

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4804813_f2_36	1381	6603	377	1134	670	8.8e-66
Protein name			Locus Name		Acc#	
			sp:ARGD_BACSU		P36839	
Description						
ACETYLORNITHINE AMINOTRANSFERASE, (ACOAT)						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
5110712_c2_114	1382	6604	413	1242	398	1.3e-37
Protein name			Locus Name		Acc#	
sensory transduction histidine kinase slr2104:protein slr2104:protein slr2104			pir:S75136		S75136	
Description						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
5131925_c1_81	1383	6605	659	1980	108	0.033
Protein name			Locus Name		Acc#	
hypothetical protein F10M10.30			pir:T04772		T04772	
Description						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
5270050_f2_33	1384	6606	160	483	254	1.1e-21
Protein name			Locus Name		Acc#	
arginine repressor			gp:BSAJ10954		AJ010954	
Description						
Bacillus stearothermophilus argR gene and partial recN gene.						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

5270302_f1_11	1385	6607	554	1665	1248	5.0e-127
---------------	------	------	-----	------	------	----------

Protein name

Locus Name

Acc#

acetyl-CoA synthetase related protein

pir:F69193

F69193

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

994002_f3_53	1386	6608	334	1005	273	1.0e-23
--------------	------	------	-----	------	-----	---------

Protein name

Locus Name

Acc#

probable malate dehydrogenase, :2-ketoacid
dehydrogenase:protein sll0891:2-ketoacid
dehydrogenase:protein sll0891

pir:S75735

S75735

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

10345327_f2_103	1387	6609	760	2283	410	7.8e-40
-----------------	------	------	-----	------	-----	---------

Protein name

Locus Name

Acc#

115K outer membrane protein precursor:SusC
protein

pir:JC6027

JC6027

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

10580052_f3_183	1388	6610	114	345		
-----------------	------	------	-----	-----	--	--

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
10662877_c1_202	1389	6611	321	966	224	1.6e-18
Protein name			Locus Name		Acc#	
putative transposase			gp:AF007429		AF007429	
Description						
Haemophilus paragallinarum IS-like putative transposase gene, complete cds.						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
10725942_c3_342	1390	6612	60	183		
Protein name			Locus Name		Acc#	
Description			NO-HIT			

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
10819681_c2_294.....	1391	6613	138	417	170	8.5e-13
Protein name			Locus Name		Acc#	
			sp:MTGA_HAEIN		P44890	
Description						
(EC 2.4.2.-) (MONOFUNCTIONAL TGASE)						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
11035088_c2_306.....	1392	6614	333	1002	1634	6.2e-168
Protein name			Locus Name		Acc#	
mobilization protein B			gp:AF118242		AF118242	
Description						
Bacteroides fragilis mobilization protein B (mobB) gene, complete cds.						

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
11832332_f1_34	1393	6615	322	969		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
11881313_c3_349.....	1394	6616	288	867	113	0.00062

Protein name

Locus Name

Acc#

transmembrane sensor

gp:AF060193

AF060193

Description

Pseudomonas aeruginosa pigACDE operon, complete sequence;hypothetical PigB (pigB) gene, complete cds.

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
12109558_c2_272.....	1395	6617	148	447	156	5.1e-11

Protein name

Locus Name

Acc#

collagen-like protein

gp:BTU67921

U67921

Description

Bacillus thuringiensis plasmid pTX14-1, MOB, REP, and collagen-likeprotein genes, complete sequence.

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
1271010_c3_387.....	1396	6618	96	291		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT</u> <u>Length</u>	<u>AA</u> <u>Length</u>	<u>Score</u>	<u>Probability</u>
-----------------	-------------	-------------	----------------------------	----------------------------	--------------	--------------------

13071943_c1_216	1397	6619	466	1401	397	7.5e-37
-----------------	------	------	-----	------	-----	---------

Protein name

Locus Name

Acc#

conserved hypothetical protein

pir:H72331

H72331

Description

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT</u> <u>Length</u>	<u>AA</u> <u>Length</u>	<u>Score</u>	<u>Probability</u>
-----------------	-------------	-------------	----------------------------	----------------------------	--------------	--------------------

13750800_c1_207	1398	6620	78	237		
-----------------	------	------	----	-----	--	--

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT</u> <u>Length</u>	<u>AA</u> <u>Length</u>	<u>Score</u>	<u>Probability</u>
-----------------	-------------	-------------	----------------------------	----------------------------	--------------	--------------------

13806517_f2_85	1399	6621	401	1206	169	1.0e-09
----------------	------	------	-----	------	-----	---------

Protein name

Locus Name

Acc#

transposase

gp:AF038866

AF038866

Description

Bacteroides fragilis transposon Tn5520 transposase (bipH) and mobilization protein BmpH (bmpH) genes, complete cds.

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT</u> <u>Length</u>	<u>AA</u> <u>Length</u>	<u>Score</u>	<u>Probability</u>
-----------------	-------------	-------------	----------------------------	----------------------------	--------------	--------------------

14469691_c1_213	1400	6622	193	582	199	7.2e-16
-----------------	------	------	-----	-----	-----	---------

Protein name

Locus Name

Acc#

RNA polymerase sigma factor SigZ-like protein

gp:AF137263

AF137263

Description

Bacteroides thetaiotaomicron 30S ribosomal protein S16-like protein, fucose gene cluster, and RNA polymerase sigma factor SigZ-like protein (sigZ) genes, complete cds.

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
14589067_f3_150	1401	6623	93	282		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
14663552_f2_7.7.....	1402	6624	296	891	116	0.00021

Protein name

Locus Name

Acc#

sp:LCRF_YERPE

P28808

Description

THERMOREGULATORY PROTEIN LCRF

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
14723156_f2_132.....	1403	6625	123	372	83	0.029

Protein name

Locus Name

Acc#

hypothetical protein aq_2087

pir:H70478

H70478

Description

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
14875302_c1_26.7.....	1404	6626	162	489		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
15659758_f1_51	1405	6627	63	192		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
15671890_c3_394	1406	6628	428	1287	191	3.9e-12

Protein name

Locus Name

Acc#

transposase

gp:AF038866

AF038866

Description

Bacteroides fragilis transposon Tn5520 transposase (bipH) and mobilization protein BmpH (bmpH) genes, complete cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
15736057_f1_29	1407	6629	523	1572	2044	2.2e-211

Protein name

Locus Name

Acc#

sp:TRA2_BACFR

Q45119

Description

TRANSPOSASE FOR INSERTION SEQUENCE ELEMENT IS21-LIKE

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
15829061_c1_200	1408	6630	478	1437	468	2.2e-44

Protein name

Locus Name

Acc#

sp:PPOX_MYXXA

P56601

Description

PROTOPORPHYRINOGEN OXIDASE, (PPO)

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
16491593_c2_279	1409	6631	157	474		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
19531438_f1_49.....	1410	6632	88	267		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
19937637_f2_111.....	1411	6633	62	189		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
2145012_f3_156.....	1412	6634	67	204	95	0.00024

Protein name

Locus Name

Acc#

hypothetical 26.8K protein

pir:JC2322

JC2322

Description

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
21683280_c3_350.....	1413	6635	527	1584		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
22459687_c3_347	1414	6636	187	564		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
22691552_c1_262.....	1415	6637	434	1305	140	2.7e-06

Protein name

Locus Name

Acc#

immunoreactive 53 kD antigen PG123

gp:AF144641

AF144641

Description

Porphyromonas gingivalis strain W50 immunoreactive 53 kD antigenPG123 gene, complete cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
22692067_c1_233.....	1416	6638	324	975	445	6.1e-42

Protein name

Locus Name

Acc#

sp:HTPX_STRGC

030795

Description

PUTATIVE HEAT SHOCK PROTEIN HTPX

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
22836061_f2_82.....	1417	6639	133	402	97	0.00018

Protein name

Locus Name

Acc#

MbpB

gp:BFU25716

U25716

Description

Bacteroides fragilis mobilization protein MbpA (mbpA), MbpB (mbpB) and MbpC (mbpC) genes, complete cds.

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
22933438_c3_397	1418	6640	322	969		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
22933438_f1_21	1419	6641	242	729		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
23495700_f2_138	1420	6642	67	204		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
23531265_f3_149	1421	6643	216	651		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
23648392_f2_86	1422	6644	431	1296	141	2.1e-06

Protein name	Locus Name	Acc#
immunoreactive 53 kD antigen PG123	gp:AF144641	AF144641

Description

Porphyromonas gingivalis strain W50 immunoreactive 53 kD antigenPG123 gene, complete cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
23679510_c3_345	1423	6645	197	594	345	2.4e-31

Protein name	Locus Name	Acc#
putative acetyltransferase	gp:SCF1	AL117322

Description

Streptomyces coelicolor cosmid F1.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24026502_f1_27.....	1424	6646	88	267		

Protein name	Locus Name	Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24259637_c3_336.....	1425	6647	690	2073	253	3.1e-37

Protein name	Locus Name	Acc#
unknown	gp:AF079317	AF079317

Description

Sphingomonas aromaticivorans plasmid pNL1, complete plasmidsequence.

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
24332035_c1_197	1426	6648	83	252		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
24407551_c2_293.....	1427	6649	231	696	427	5.0e-40

Protein name

Locus Name

Acc#

immunogenic 23 kDa lipoprotein PG3

gp:AF145799

AF145799

Description

Porphyromonas gingivalis strain W50 immunogenic 23 kDa lipoproteinPG3 gene, complete cds.

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
24415757_c3_348.....	1428	6650	424	1275		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
24641061_f2_96.....	1429	6651	316	951	147	9.5e-08

Protein name

Locus Name

Acc#

vr1E protein

pir:T17384

T17384

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24642137_f1_62	1430	6652	401	1206	176	2.9e-11

Protein name	Locus Name	Acc#
putative outer membrane porin	gp:AF030977	

Description

Vibrio cholerae glutamyl tRNA synthetase (gltX) gene, partial cds; putative outer membrane porin (ompA), unknown protein, vibriobactin receptor precursor (viuA), and ViuB protein (viuB) genes, complete cds; and VibF (vibF) gene, partial cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24642212_f2_110	1431	6653	301	906	631	1.2e-61

Protein name	Locus Name	Acc#
	sp:YBFH_BACSU	031448

Description

HYPOTHETICAL 33.8 KD PROTEIN IN GLPT-PURT INTERGENIC REGION

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24647826_f1_28	1432	6654	77	234		

Protein name	Locus Name	Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24725380_c1_245	1433	6655	186	561	593	1.3e-57

Protein name	Locus Name	Acc#

mobilization protein A

gp:AF118241

AF118241

Description

Bacteroides fragilis mobilization protein A (mobA) gene, complete cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24726592_c3_355	1434	6656	152	459	339	1.0e-30

Protein name

Locus Name

Acc#

sp:MTGA_ECOLI

P46022

Description

(EC 2.4.2.-) (MONOFUNCTIONAL TGASE)

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24803426_c1_210	1435	6657	204	615	120	1.9e-06

Protein name

Locus Name

Acc#

hypothetical protein MTH847

pir:A69213

A69213

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24847551_c2_302	1436	6658	92	279		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
2538277_c2_282	1437	6659	105	318	120	1.7e-07

Protein name

Locus Name

Acc#

hypothetical protein ydaT

pir:C69770

C69770

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
25511052_c1_232	1438	6660	206	621	401	2.8e-37

Protein name	Locus Name	Acc#
LemA	gp:LMU66186	U66186

Description

Listeria monocytogenes LemA (lemA) gene, complete cds, and LemB(lemB) gene, partial cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
25527053_f1_22	1439	6661	436	1311		

Protein name	Locus Name	Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
26354518_f2_95.....	1440	6662	73	222		

Protein name	Locus Name	Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
2767137_c1_266.....	1441	6663	887	2664	128	0.00031

Protein name	Locus Name	Acc#

hypothetical protein H02F09.3	pir:T33369	T33369
-------------------------------	------------	--------

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
2792942_f2_70	1442	6664	87	264		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
29713458_f2_83	1443	6665	67	204		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
31923438_c2_291	1444	6666	61	186		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
32213312_f3_163	1445	6667	171	516	237	6.8e-20

Protein name

Locus Name

Acc#

putative ECF sigma factor RpoE1

gp:AF049107

AF049107

Description

Myxococcus xanthus response regulator FrzZ (frzZ) gene, partialcds; alanine dehydrogenase (aldA), putative ECF sigma factor RpoE1(rpoE1), and response regulator homolog (frzS) genes, complete cds;and unknown genes.

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
33214538_f3_155	1446	6668	744	2235		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
33486716_c2_276.....	1447	6669	497	1494	793	8.2e-79

Protein name

Locus Name

Acc#

sp:HEMN_AQUAE

067886

Description

OXYGEN-INDEPENDENT COPROPORPHYRINOGEN II

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
33632692_c1_229.....	1448	6670	71	216		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
34181502_f2_122.....	1449	6671	426	1281	134	6.9e-06

Protein name

Locus Name

Acc#

probable carboxy-terminal proteinase, D1

pir:T05975

T05975

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
35302_f1_52	1450	6672	203	612		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
3.942130...c1...264.....	1451	6673	388	1167	86	0.0055

Protein name

Locus Name

Acc#

integrase

gp:HIVU69223

U69223

Description

HIV-1 strain CMR273 from Cameroon integrase (pol) gene, partial cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
3.954762...f3...187.....	1452	6674	291	876	111	6.4e-06

Protein name

Locus Name

Acc#

transcription regulator homolog: hypothetical
137 protein

pir:PC4110

PC4110

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4114055...c3...337.....	1453	6675	415	1248	288	5.0e-24

Protein name

Locus Name

Acc#

hypothetical protein

gp:AF149851

AF149851

Description

Pseudomonas sp. KC hypothetical proteins, methallothionein-like protein, MoeB-like protein, putative proteins, hypothetical protein, putative oxidoreductase, and putative AMP ligase (entE) genes, complete cds; and putative receptor gene, partial cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4720187_f2_99	1454	6676	283	852	954	7.1e-96

Protein name

Locus Name

Acc#

sp:ISTB_BACFR

Q45120

Description

INSERTION SEQUENCE IS21-LIKE PUTATIVE ATP-BINDING PROTEIN

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4822751_f2_101	1455	6677	594	1785	370	1.8e-48

Protein name

Locus Name

Acc#

oxaloacetate decarboxylase, subunit alpha
(oadA) homolog

pir:C69406

C69406

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4864702_f2_124	1456	6678	152	459		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4879635_f1_61	1457	6679	772	2319	214	4.6e-14

Protein name

Locus Name

Acc#

collagen

gp:AB008933

AB008933

Description

Hydra vulgaris HT2 mRNA for collagen, partial cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
5177157_f2_88	1458	6680	182	549		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
5285692_c2_281	1459	6681	86	261		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
5321932_f1_53	1460	6682	234	705		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
564037_c3_346	1461	6683	282	849	229	4.8e-19

Protein name

Locus Name

Acc#

hypothetical protein

pir:B72308

B72308

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
5970252_c2_316	1462	6684	61	186		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
6025010_c3_333	1463	6685	74	225		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
6046907_f1_35	1464	6686	636	1911		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
6775438_c3_401	1465	6687	124	375		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
785322_f1_64	1466	6688	584	1752	220	2.7e-20

Protein name

Locus Name

Acc#

sp:FECA_ECOLI

P13036

Description

IRON(III) DICITRATE TRANSPORT PROTEIN FECA PRECURSOR

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
15808290_c1_33	1467	6689	61	186	99	2.8e-05
<u>Protein name</u>			<u>Locus Name</u>			<u>Acc#</u>
glycine-rich protein (clone w10-1)			pir:S14982			S14982
<u>Description</u>						

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
19659503_c1_32	1468	6690	383	1152		
<u>Protein name</u>			<u>Locus Name</u>			<u>Acc#</u>
<u>Description</u>						

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
26367141_c1_30	1469	6691	292	879	130	3.5e-05
<u>Protein name</u>			<u>Locus Name</u>			<u>Acc#</u>
membrane glycoprotein			gp:D88733			D88733
<u>Description</u>						

Equine herpesvirus 1 DNA for membrane glycoprotein, complete cds.

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
26425336_c2_38	1470	6692	250	753		
<u>Protein name</u>			<u>Locus Name</u>			<u>Acc#</u>
<u>Description</u>						

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
34666302_c3_43	1471	6693	495	1488	213	1.3e-20

Protein name

Locus Name

Acc#

immunoreactive 53 kD antigen PG123

gp:AF144641

AF144641

Description

Porphyromonas gingivalis strain W50 immunoreactive 53 kD antigenPG123 gene, complete cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
807033_c1_29	1472	6694	112	339		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
11063391_c2_40.....	1473	6695	344	1035	452	1.1e-42

Protein name

Locus Name

Acc#

sp:YF23_HAEIN

P44243

Description

HYPOTHETICAL PROTEIN HI1523

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
14348958_f2_14.....	1474	6696	87	264		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
14650012_f3_26	1475	6697	145	435	73	0.022

Protein name	Locus Name	Acc#
glucosidase II beta-subunit	gp:AF066061	AF066061

Description

Mus musculus glucosidase II beta-subunit gene, alternatively spliced products, partial cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
15835261_c1_28	1476	6698	116	351		

Protein name	Locus Name	Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
23472832_c1_31	1477	6699	366	1101	454	6.8e-43

Protein name	Locus Name	Acc#
	sp:CBH_CLOPE	P54965

Description

HYDROLASE) (CBAH) (BILE SALT HYDROLASE)

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24337962_f1_6	1478	6700	62	189		

Protein name	Locus Name	Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24798401_f1_2	1479	6701	141	426		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
30355305_f3_24	1480	6702	271	816	409	4.0e-38

Protein name

Locus Name

Acc#

sp:SOJ_BACSU

P37522

Description

SOJ PROTEIN

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
33882837_f3_25	1481	6703	97	294	80	0.024

Protein name

Locus Name

Acc#

hypothetical protein F20D10.230

pir:T05638

T05638

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4407575_f2_12	1482	6704	522	1569	115	2.8e-10

Protein name

Locus Name

Acc#

endo-xylanase homolog PCZA361.14

pir:T17480

T17480

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
5869167_f1_7	1483	6705	73	222		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

6838437_f1_1	1484	6706	422	1269	100	0.0024
--------------	------	------	-----	------	-----	--------

Protein name	Locus Name	Acc#
--------------	------------	------

outer membrane protein	gp:BNROMPB	L77614
------------------------	------------	--------

Description

Bacteroides thetaiotaomicron outer membrane protein (susD) gene, complete cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

10547256_f2_6	1485	6707	81	246		
---------------	------	------	----	-----	--	--

Protein name	Locus Name	Acc#
--------------	------------	------

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

11126552_c2_14	1486	6708	152	459	95	7.5e-05
----------------	------	------	-----	-----	----	---------

Protein name	Locus Name	Acc#
--------------	------------	------

hypothetical protein aq_1018	pir:H70387	H70387
------------------------------	------------	--------

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

2117841_f1_1	1487	6709	749	2250		
--------------	------	------	-----	------	--	--

Protein name	Locus Name	Acc#
--------------	------------	------

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
26682828_c3_19	1488	6710	136	411	93	0.023
Protein name			Locus Name		Acc#	
surface exclusion protein sepl precursor			pir:S72375		S72375	
Description						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
31417187_c1_10	1489	6711	261	786		
Protein name			Locus Name		Acc#	
Description			NO-HIT			

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
6035687_c2_18	1490	6712	265	795	1378	8.3e-141
Protein name			Locus Name		Acc#	
fructanase			pir:A36915		A36915	
Description						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
85912_c1_11	1491	6713	790	2373	2229	5.5e-231
Protein name			Locus Name		Acc#	
Description			gp:BNRSCRL		M83774	
Bacteroides fragilis levanase (scrl) gene, complete cds.						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
14333277_c1_10	1492	6714	127	384		
Protein name			Locus Name		Acc#	
Description			NO-HIT			

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
15735882_f3_5	1493	6715	400	1203	696	1.5e-68

<u>Protein name</u>	<u>Locus Name</u>	<u>Acc#</u>
renin-binding protein-related protein:protein slr1975:protein slr1975	pir:S75649	S75649

Description

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
26819566_f3_4	1494	6716	73	222		

<u>Protein name</u>	<u>Locus Name</u>	<u>Acc#</u>

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
31381_f3_6	1495	6717	417	1254	280	8.7e-23

<u>Protein name</u>	<u>Locus Name</u>	<u>Acc#</u>

hexuronate transporter homolog yjmg	pir:A69853	A69853
-------------------------------------	------------	--------

Description

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
3956707_f1_1	1496	6718	149	450	126	1.4e-07

<u>Protein name</u>	<u>Locus Name</u>	<u>Acc#</u>

N-acetylneuraminate lyase	gp:CPNANA	Y12876
---------------------------	-----------	--------

Description

C.perfringens gene encoding N-acetylneuraminate lyase and twopartial open reading frames.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
5117337_f2_3	1497	6719	150	453	201	4.4e-16

Protein name

Locus Name

Acc#

sp:YHCH_HAEIN

P44583

Description

HYPOTHETICAL PROTEIN HI0227

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
781932_f3_7	1498	6720	698	2094	543	3.1e-53

Protein name

Locus Name

Acc#

115K outer membrane protein precursor:SusC protein

pir:JC6027

JC6027

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
11209542_c3_31	1499	6721	73	222		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24023442_c3_32	1500	6722	329	990	630	1.5e-61

Protein name

Locus Name

Acc#

metabolite transporter homolog yfnA

pir:D69814

D69814

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
25433212_f1_2	1501	6723	116	351	193	5.1e-14
Protein name			Locus Name		Acc#	
alpha-N-acetylglucosaminidase			gp:NTA18209		Y18209	
Description						
Nicotiana tabacum mRNA for alpha-N-acetylglucosaminidase.						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
26681533_c2_26	1502	6724	202	609	327	3.5e-29
Protein name			Locus Name		Acc#	
probable cationic amino acid transporter			pir:T34694		T34694	
Description						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
29931309_c1_12	1503	6725	432	1299	195	1.9e-12
Protein name			Locus Name		Acc#	
immunoreactive 52kD antigen PG41			gp:AF175716		AF175716	
Description						
Porphyromonas gingivalis strain W50 immunoreactive 52kD antigen PG41 gene, complete cds.						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
30516442_f1_1	1504	6726	446	1341	618	2.9e-60
Protein name			Locus Name		Acc#	
			sp:ANAG_HUMAN		P54802	
Description						
GLUCOSAMINIDASE) (NAG)						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
10580062_c2_93	1505	6727	301	903	279	8.0e-24
<u>Protein name</u>			<u>Locus Name</u>		<u>Acc#</u>	
60kDa protein			gp:AB004560		AB004560	
<u>Description</u>						
Porphyromonas gingivalis DNA for 60kDa protein, complete cds.						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
13175950_c2_91	1506	6728	72	219		
<u>Protein name</u>			<u>Locus Name</u>		<u>Acc#</u>	
<u>Description</u>						
NO-HIT						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
13869000_c2_78	1507	6729	999	3000	886	1.6e-109
<u>Protein name</u>			<u>Locus Name</u>		<u>Acc#</u>	
115K outer membrane protein precursor:SusC protein			pir:JC6027		JC6027	
<u>Description</u>						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
14347666_f2_26	1508	6730	68	207		
<u>Protein name</u>			<u>Locus Name</u>		<u>Acc#</u>	
<u>Description</u>						
NO-HIT						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
I4647808_f2_20	1509	6731	287	864	126	4.3e-06

Protein name

Locus Name

Acc#

sp:YDIP_ECOLI

P77402

Description

HYPOTHETICAL TRANSCRIPTIONAL REGULATOR IN AROD-PPS INTERGENIC REGION

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
I4660892_c1_74	1510	6732	207	624		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
I5835062_f1_18	1511	6733	96	291	78	0.0048

Protein name

Locus Name

Acc#

hypothetical protein c04005

pir:S75372

S75372

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
20007812_c3_102	1512	6734	467	1404	195	1.7e-12

Protein name

Locus Name

Acc#

transposase

gp:AF038866

AF038866

Description

Bacteroides fragilis transposon Tn5520 transposase (bipH) and mobilization protein BmpH (bmpH) genes, complete cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
22683287_f3_56	1513	6735	486	1461		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24492177_c2_88	1514	6736	342	1029		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24895165_f1_13	1515	6737	65	198	47	0.029

Protein name

Locus Name

Acc#

hypothetical protein T11B7.2

pir:T24826

T24826

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
26771041_c3_94	1516	6738	260	783	417	5.7e-39

Protein name

Locus Name

Acc#

hypothetical protein C33G8.2

pir:T34137

T34137

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
33772811_f3_47	1517	6739	109	330		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
34510418_c1_63	1518	6740	259	780	424	1.0e-39
Protein name			Locus Name		Acc#	
hypothetical protein F36H12.3			pir:T33457		T33457	
Description						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
35734500_c2_79	1519	6741	512	1539	202	7.9e-13
Protein name			Locus Name		Acc#	
unknown			gp:U96771		U96771	
Description						
Prevotella bryantii putative polygalacturonase, B-1,4-endoglucanase, and mannanase genes, complete cds; and unknown genes.						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
36198958_c2_76	1520	6742	278	837	430	2.4e-40
Protein name			Locus Name		Acc#	
hypothetical protein C33G8.2			pir:T34137		T34137	
Description						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
36574092_c2_80	1521	6743	430	1293	545	3.1e-59
Protein name			Locus Name		Acc#	
			sp:YBDN_ECOLI		P77216	
Description						
HYPOTHETICAL 47.8 KD PROTEIN IN CSTA-DSBG INTERGENIC REGION						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
397175_c2_77	1522	6744	176	531		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4062906_c3_101	1523	6745	785	2358	130	1.3e-08

Protein name

Locus Name

Acc#

sp:FYUA_YEREN

P46360

Description

PESTICIN RECEPTOR PRECURSOR (IRPC) (IPR65)

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4969091_c3_97	1524	6746	138	417		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
5900377_c1_66	1525	6747	177	534	351	5.6e-32

Protein name

Locus Name

Acc#

sp:YBDM_ECOLI

P77174

Description

HYPOTHETICAL 23.9 KD PROTEIN IN CSTA-DSBG INTERGENIC REGION

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
6485055_c2_89	1526	6748	92	279		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
103287_f3_99	1527	6749	501	1506	479	1.5e-45

Protein name

Locus Name

Acc#

immunoreactive 51kD antigen PG52

gp:AF175719

AF175719

Description

Porphyromonas gingivalis strain W50 immunoreactive 51kD antigenPG52 gene, complete cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
11976562_c3_189	1528	6750	61	186		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
12304718_f1_2	1529	6751	518	1557	139	3.1e-13

Protein name

Locus Name

Acc#

sp:BGAL_THEET

P77989

Description

BETA-GALACTOSIDASE, (LACTASE)

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
13704552_c1_129	1530	6752	496	1491	1403	1.9e-143
<u>Protein name</u>			<u>Locus Name</u>			<u>Acc#</u>
			sp:6PGD_TREPA			083351

Description

6-PHOSPHOGLUCONATE DEHYDROGENASE, DECARBOXYLATING,

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
13758530_c3_190	1531	6753	136	411		
<u>Protein name</u>			<u>Locus Name</u>			<u>Acc#</u>

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
13907312_c3_186.....	1532	6754	74	225	77	0.0096
<u>Protein name</u>			<u>Locus Name</u>			<u>Acc#</u>
putative signal transduction protein GarA			gp:AF173844			AF173844

Description

Mycobacterium smegmatis garA-containing gene cluster, partial sequence.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
13962757_c2_175.....	1533	6755	383	1152	350	7.2e-32
<u>Protein name</u>			<u>Locus Name</u>			<u>Acc#</u>
cytochrome d oxidase subunit II			gp:AF001503			AF001503

Description

Salmonella typhimurium cytochrome d oxidase subunit I (cydA) and cytochrome d oxidase subunit II (cydB) genes, complete cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
1444627_c1_132	1534	6756	62	189	58	0.039
Protein name			Locus Name		Acc#	
ribosomal protein S5			gp:U87145		U87145	
Description						
Toxoplasma gondii chloroplast, complete genome.						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
16441305_f3_103	1535	6757	236	711	244	1.2e-20
Protein name			Locus Name		Acc#	
hypothetical protein b2381			pir:B65012		B65012	
Description						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
165117156_f3_101	1536	6758	450	1353	717	9.2e-71
Protein name			Locus Name		Acc#	
hypothetical protein			pir:S76946		S76946	
Description						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
194128_c1_130	1537	6759	84	255	63	0.0078
Protein name			Locus Name		Acc#	
p20-CGGBP			gp:HSCGGBP		AJ000258	
Description						
Homo sapiens trinucleotide repeat 5-d(CGG)n-3ds binding proteinp20-CGGBP.						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
19687836_f3_87	1538	6760	429	1290	945	6.4e-95

Protein name

Locus Name

Acc#

sp:YCAJ_HAEIN

P45262

Description

HYPOTHETICAL PROTEIN H11590

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
2068766_c1_143	1539	6761	525	1578	1098	3.9e-111

Protein name

Locus Name

Acc#

sp:CYDA_AZOVI

Q09049

Description

CYTOCHROME D UBIQUINOL OXIDASE SUBUNIT I,

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
20735878_c2_163	1540	6762	142	429		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
2204377_f2_70	1541	6763	446	1341	911	2.6e-91

Protein name

Locus Name

Acc#

RumB (R391)

gp:XXU13633

U13633

Description

IncJ plasmid R391 rumA(R391) and rumB(R391) genes, complete cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
22273312_c1_131	1542	6764	287	864	271	1.7e-23

Protein name	Locus Name	Acc#
urea transport protein	gp:AF167577	AF167577

Description

Actinobacillus pleuropneumoniae transcriptional regulator (apuR) gene, partial cds; and putative periplasmic binding protein (cbiK), putative cytoplasmic membrane protein (cbiL), cobalt membranetransport protein homolog (cbiM), cobalt membrane transport proteinhomolog (cbiQ), cobalt transport ATP-binding protein homolog (cbiO), and urea transport protein

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
22535925_c2_176	1543	6765	359	1080	184	2.2e-12

Protein name	Locus Name	Acc#
molybdate metabolism regulator	pir:B64979	B64979

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
22941306_f2_62	1544	6766	258	777	651	9.1e-64

Protein name	Locus Name	Acc#
ABC transporter, ATP-binding protein	pir:H72385	H72385

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
23985880_f1_26	1545	6767	105	318		

Protein name	Locus Name	Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24100265_c3_188	1546	6768	505	1518	1300	1.5e-132

Protein name

Locus Name

Acc#

sp:G6PD_ACTAC

P77809

Description

GLUCOSE-6-PHOSPHATE 1-DEHYDROGENASE, (G6PD)

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24229800_f3_98	1547	6769	79	240		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24647180_c3_191	1548	6770	690	2073	101	0.0017

Protein name

Locus Name

Acc#

hypothetical protein MTH357

pir:A69146

A69146

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
25517013_f1_1	1549	6771	297	894	392	3.8e-35

Protein name

Locus Name

Acc#

putative secreted beta-galactosidase

gp:SCF81

AL133171

Description

Streptomyces coelicolor cosmid F81.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
25667675_c3_205	1550	6772	341	1026		

Protein name Locus Name Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
25970016_f3_100	1551	6773	411	1236	326	2.5e-29

Protein name Locus Name Acc#

probable membrane protein b0878 pir:F64826 F64826

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
31769537_f1_24	1552	6774	206	621	222	2.0e-17

Protein name Locus Name Acc#

sp:YEHU_ECOLI

Description

HYPOTHETICAL 62.1 KD PROTEIN IN MOLR-BGLX INTERGENIC REGION PRECURSOR

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
31886308_c2_159	1553	6775	259	780	387	8.6e-36

Protein name Locus Name Acc#

probable glucose-6-phosphate 1-dehydrogenase pir:C71319 C71319

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
3257635_c1_134	1554	6776	426	1281		

Protein name Locus Name Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
34382687_c3_193	1555	6777	419	1260		

Protein name Locus Name Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
3945257_f3_102	1556	6778	158	477		

Protein name Locus Name Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4103890_c1_147	1557	6779	515	1548	145	9.4e-08

Protein name Locus Name Acc#

conserved hypothetical protein AF0444 pir:D69305 D69305

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4737662_f1_39	1558	6780	395	1188	578	4.9e-56

Protein name Locus Name Acc#

probable glutamate/ aspartate transporter pir:G71309 G71309

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
5117762_f3_106	1559	6781	149	450	304	5.4e-27

Protein name Locus Name Acc#

RumA(R391) gp:XXU13633 U13633

Description

Incl plasmid R391 rumA(R391) and rumB(R391) genes, complete cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
5275250_f2_47	1560	6782	319	960	554	1.7e-53

Protein name

Locus Name

Acc#

sp:DHGY_METEX

Q59516

Description

REDUCTASE) (HPR-A)

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
7287787_c1_133	1561	6783	267	804		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
9770137_c1_145.....	1562	6784	287	864	109	5.1e-09

Protein name

Locus Name

Acc#

gp:AB016260

Description

Agrobacterium tumefaciens plasmid pTi-SAKURA, complete sequence.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
9922057_c3_206.....	1563	6785	432	1299	293	8.0e-26

Protein name

Locus Name

Acc#

coproporphyrinogen oxidase, III,
oxygen-independent hemN

pir:B69640

B69640

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
13869003_f3_21	1564	6786	535	1608	123	0.0010
Protein name			Locus Name		Acc#	
glycoprotein Vp260-like protein A18L			pir:T17508		T17508	
Description						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
23864381_c2_38	1565	6787	474	1425	747	6.1e-74
Protein name			Locus Name		Acc#	
metabolite transport protein homolog ywtG			pir:E70070		E70070	
Description						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
25632943_c1_29	1566	6788	184	555		
Protein name			Locus Name		Acc#	
Description			NO-HIT			

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
26364040_f1_6	1567	6789	61	186		
Protein name			Locus Name		Acc#	
Description			NO-HIT			

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
33400260_f3_23	1568	6790	503	1512	124	0.00028
Protein name			Locus Name		Acc#	
STARP antigen			gp:PFSTARP		Z26314	
Description						
P.falciparum gene for STARP antigen.						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
791406_c3_44	1569	6791	70	213	77	0.026

Protein name

Locus Name

Acc#

sp:ATP6_ACACA

Q37385

Description

ATP SYNTHASE A CHAIN, (PROTEIN 6)

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
9862501_c3_41	1570	6792	109	330		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
14664052_f3_7	1571	6793	205	618	545	1.6e-52

Protein name

Locus Name

Acc#

gp:PGPGAAGEN

X95938

Description

P.gingivalis rnhB & pgaA genes & orfs 150, 197, 202 & 199.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
34406512_f1_4	1572	6794	311	936	663	4.9e-65

Protein name

Locus Name

Acc#

2,3-bisphosphoglycerate-independent

gp:AF120090

AF120090

Description

Bacillus megaterium 2,3-bisphosphoglycerate-independent phosphoglycerate mutase (pgm) gene, complete cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
36135311_c1_9	1573	6795	315	948	444	7.8e-42
Protein name				Locus Name		Acc#
probable transport protein				pir:A75272		A75272
Description						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
36330175_f1_5	1574	6796	62	186		
Protein name				Locus Name		Acc#
Description						

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
15039156_f3_6	1575	6797	192	579	398	4.7e-36
Protein name				Locus Name		Acc#
putative large secreted protein				gp:SCF12		AL117669
Description						

Streptomyces coelicolor cosmid F12.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
15117192_f2_5	1576	6798	89	270	79	0.042
Protein name				Locus Name		Acc#
				gp:PFMAL3P7		
Description						

Plasmodium falciparum MAL3P7, complete sequence.

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
24343756_c1_7	1577	6799	62	189		

Protein name Locus Name Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
4860650_f1_3	1578	6800	95	288	75	0.0099

Protein name Locus Name Acc#

ct602 hypothetical protein pir:F72036 F72036

Description

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
10660763_c3_339	1579	6801	387	1164		

Protein name Locus Name Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
1176302_f1_16	1580	6802	799	2400	150	3.0e-14

Protein name Locus Name Acc#

putative TonB-dependent outer membrane receptor gp:AF048749 AF048749

Description

Bacteroides fragilis capsular polysaccharide biosynthesis operon, complete sequence.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
12117076_c1_211	1581	6803	67	204		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
12506402_f1_15	1582	6804	954	2865	249	4.1e-17

Protein name

Locus Name

Acc#

putative histidine protein kinase

gp:REU82564

U82564

Description

hydrogenase-like protein small subunit(hoxB) gene, hydrogenase-like protein large subunit (hoxC) gene, and putative histidine protein kinase (hoxJ) gene, complete cds, and nickel permease (hoxN) gene, partial cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
12540880_c3_343	1583	6805	343	1032	361	4.9e-33

Protein name

Locus Name

Acc#

capsular polysaccharide biosynthesis homolog yveT

pir:A70037

A70037

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
12714062_c3_354	1584	6806	76	231		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
1290933_f3_136	1585	6807	143	432	165	2.9e-12
Protein name			Locus Name		Acc#	
hypothetical protein slr1861			pir:S77097		S77097	
Description						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
12923260_c2_271	1586	6808	516	1551	207	1.6e-13
Protein name			Locus Name		Acc#	
putative flippase			gp:AF125164		AF125164	
Description						
Bacteroides fragilis 638R polysaccharide B (PS B2) biosynthesis locus, complete sequence; and unknown genes.						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
14070180_c1_202	1587	6809	150	453	175	2.5e-13
Protein name			Locus Name		Acc#	
hypothetical protein 1			pir:S28678		S28678	
Description						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
1448442_c1_203	1588	6810	354	1065	599	2.9e-58
Protein name			Locus Name		Acc#	
mannose-1-phosphate guanylyltransferase			pir:H72303		H72303	
Description						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
14567135_c1_201	1589	6811	369	1110	118	0.00040

Protein name

Locus Name

Acc#

immunoreactive 43kD antigen PG32

gp:AF175714

AF175714

Description

Porphyromonas gingivalis strain W50 immunoreactive 43kD antigenPG32 gene, complete cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
14658342_f2_65	1590	6812	583	1752	139	2.5e-06

Protein name

Locus Name

Acc#

hypothetical protein SPAC17G6.19c

pir:T37851

T37851

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
14726512_f1_11	1591	6813	105	318	161	7.6e-12

Protein name

Locus Name

Acc#

hypothetical protein slr1856

pir:S77093

S77093

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
15628390_f1_9	1592	6814	647	1944	979	1.6e-98

Protein name

Locus Name

Acc#

sp:CAPD_STAAU

P39853

Description

CAPD PROTEIN

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
15797007_c2_274	1593	6815	383	1152	322	6.6e-29

Protein name	Locus Name	Acc#
Cps1K	gp:AF155804	AF155804

Description

Streptococcus suis strain 6555 Cps1E (cps1E) gene, partial cds; Cps2F (cps2F), Cps1G (cps1G), Cps1H (cps1H), Cps1I (cps1I), and Cps1J (cps1J) genes, complete cds; and Cps1K (cps1K) gene, partial cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
15822807_f1_2	1594	6816	549	1650		

Protein name	Locus Name	Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
171902_f1_31.....	1595	6817	67	204	49	0.037

Protein name	Locus Name	Acc#

probable RNA-directed DNA polymerase, :reverse transcriptase

pir:S20016	S20016
------------	--------

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
19710937_f2_123.....	1596	6818	490	1473		

Protein name	Locus Name	Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
19728412_c3_391	1597	6819	459	1380	472	9.2e-54

Protein name	Locus Name	Acc#
folypolyglutamate synthase/dihydrofolate synthase	pir:D72411	D72411

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
2032137_c1_213	1598	6820	88	267		

Protein name	Locus Name	Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
21651557_c3_338	1599	6821	354	1065	128	2.2e-05

Protein name	Locus Name	Acc#
hypothetical protein RP338	pir:D71690	D71690

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
23455077_c3_348	1600	6822	421	1266		

Protein name	Locus Name	Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

23594641_c1_195	1601	6823	250	753	316	2.9e-28
-----------------	------	------	-----	-----	-----	---------

Protein name	Locus Name	Acc#
--------------	------------	------

putative UDP-N-acetyl-D-mannosamine transferase	gp:SPU09239	U09239
---	-------------	--------

Description

Streptococcus pneumoniae type 19F capsular polysaccharide biosynthesis operon, (cps19fABCDEFGHIJKLMNO) genes, complete cds, and aliA gene, partial cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

23632802_f3_146	1602	6824	270	813	412	1.9e-38
-----------------	------	------	-----	-----	-----	---------

Protein name	Locus Name	Acc#
--------------	------------	------

gp:AB008550	AB008550
-------------	----------

Description

Pseudomonas aeruginosa phage phi CTX, complete genome sequence.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

2377092_f3_153	1603	6825	64	195	219	1.3e-17
----------------	------	------	----	-----	-----	---------

Protein name	Locus Name	Acc#
--------------	------------	------

putative aminotransferase	gp:AF125164	AF125164
---------------------------	-------------	----------

Description

Bacteroides fragilis 638R polysaccharide B (PS B2) biosynthesis locus, complete sequence; and unknown genes.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

24038512_f3_134	1604	6826	251	756	375	1.6e-34
-----------------	------	------	-----	-----	-----	---------

Protein name	Locus Name	Acc#
--------------	------------	------

sp:YACO_BACSU	Q06753
---------------	--------

Description

HYPOTHETICAL TRNA/RRNA METHYLTRANSFERASE YACO,

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24412537_f1_20	1605	6827	276	831		

Protein name Locus Name Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24413887_c1_200	1606	6828	178	537	72	0.048

Protein name Locus Name Acc#

sp:Y235_METJA Q57687

Description

HYPOTHETICAL PROTEIN MJ0235

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24417550_f2_120	1607	6829	63	192	71	0.026

Protein name Locus Name Acc#

sp:FLIT_BACSU P39740

Description

FLAGELLAR PROTEIN FLIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24475937_c3_328	1608	6830	80	243		

Protein name Locus Name Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
24506692_c1_243	1609	6831	201	606		

Protein name Locus Name Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
24642837_f3_145	1610	6832	114	345		

Protein name Locus Name Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
24647535_f1_59	1611	6833	323	972		

Protein name Locus Name Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
24651502_c1_218	1612	6834	81	246		

Protein name Locus Name Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24694187_c1_198	1613	6835	449	1350	128	0.00060

Protein name	Locus Name	Acc#
lacunin	gp:AF078161	AF078161

Description

Manduca sexta lacunin mRNA, complete cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24864003_f3_131	1614	6836	403	1212	842	5.2e-84

Protein name	Locus Name	Acc#
pantothenate metabolism flavoprotein dfp homolog yloI:probable aspartate 1-decarboxylase activase	pir:D69878	D69878

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
25517305_c3_352	1615	6837	455	1368		

Protein name	Locus Name	Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
25578390_f1_30	1616	6838	418	1257	862	4.0e-86

Protein name	Locus Name	Acc#

methyImalonyl-CoA decarboxylase, beta-subunit	gp:PMAJ2015	AJ002015
---	-------------	----------

Description

Propionigenium modestum mmdD, mmdC, mmdB genes and partial mmdAgene.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
25625438_c1_190	1617	6839	398	1197	109	0.0069

Protein name	Locus Name	Acc#
transmembrane protein	gp:YSCPTM	L11895

Description

Saccharomyces cerevisiae putative transmembrane protein (PTM1) gene, complete cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
26210302_f1_10	1618	6840	393	1182	252	3.3e-36

Protein name	Locus Name	Acc#
sensory transduction system regulatory protein slr1983:protein slr1983:protein slr1983	pir:S75664	S75664

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
26366312_c2_287.....	1619	6841	159	480		

Protein name	Locus Name	Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
26369000_f2_81.....	1620	6842	134	405	91	0.0067

Protein name	Locus Name	Acc#
positive regulator for virulence factors	gp:CLOORF1	D14877

Description

Clostridium perfringens virR gene for positive regulator for virulence factors, complete cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

26595260_f2_82	1621	6843	206	621	120	5.9e-06
----------------	------	------	-----	-----	-----	---------

Protein name	Locus Name	Acc#
--------------	------------	------

hypothetical protein AF0417	pir:A69302	A69302
-----------------------------	------------	--------

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

26687791_f3_144	1622	6844	193	582		
-----------------	------	------	-----	-----	--	--

Protein name	Locus Name	Acc#
--------------	------------	------

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

26815891_c1_187	1623	6845	190	573	213	2.4e-17
-----------------	------	------	-----	-----	-----	---------

Protein name	Locus Name	Acc#
--------------	------------	------

unknown	gp:AF048749	AF048749
---------	-------------	----------

Description

Bacteroides fragilis capsular polysaccharide biosynthesis operon, complete sequence.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

2767217_f2_89	1624	6846	549	1650	428	3.3e-39
---------------	------	------	-----	------	-----	---------

Protein name	Locus Name	Acc#
--------------	------------	------

2', 3'-cyclic nucleotide 2'-phosphodiesterase	gp:AB028630	AB028630
---	-------------	----------

Description

Clostridium perfringens hyp27, bach, ptp, cpd genes for hypothetical protein, bacterial hemoglobin, protein-tyrosine phosphatase, 2', 3'-cyclic nucleotide 2'-phosphodiesterase, partial and complete cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
2848255_f2_124	1625	6847	75	228	106	1.1e-05

Protein name

Locus Name

Acc#

GlyA

gp:AF136495

AF136495

Description

Campylobacter lari GlyA (glyA) gene, partial cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
2931557_f1_4	1626	6848	258	777	217	8.9e-18

Protein name

Locus Name

Acc#

probable DNA pol III epsilon chain

pir:B71536

B71536

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
29563916_c3_345.....	1627	6849	368	1107	366	1.4e-33

Protein name

Locus Name

Acc#

galactosyl transferase

gp:SPN239004

AJ239004

Description

Streptococcus pneumoniae type 8 capsular gene cluster.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
31423265_c2_276.....	1628	6850	378	1137		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
31849128_f3_133	1629	6851	298	897	420	2.7e-39
Protein name			Locus Name		Acc#	
DNA repair protein			pir:A75391		A75391	
Description						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
32040875_f3_132	1630	6852	263	792	394	1.6e-36
Protein name			Locus Name		Acc#	
			sp:REC_N_ECOLI			
Description						

DNA REPAIR PROTEIN REC_N (RECOMBINATION PROTEIN N)

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
32681627_f1_3	1631	6853	151	456	128	2.4e-08
Protein name			Locus Name		Acc#	
			sp:DP3B_VIBHA		P52620	
Description						

DNA POLYMERASE III, BETA CHAIN, (FRAGMENT)

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
3322152_c1_199	1632	6854	192	579		
Protein name			Locus Name		Acc#	
Description						

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
33788882_c1_212	1633	6855	228	687	395	1.2e-36

Protein name	Locus Name	Acc#
conserved hypothetical protein aq_274	pir:C70325	C70325
<u>Description</u>		

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
34070311_c1_191	1634	6856	347	1044	132	7.0e-06

Protein name	Locus Name	Acc#
transmembrane protein	gp:SPAJ6986	AJ006986
<u>Description</u>		

Streptococcus pneumoniae type 33F DNA, capsular gene cluster.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
35784765_f2_61	1635	6857	316	951	593	1.3e-57

Protein name	Locus Name	Acc#
UDP-N-acetylenolpyruvoylglucosamine reductase	gp:BPE238308	AJ238308
<u>Description</u>		

Bordetella pertussis partial gene for putative thioesterase, tRNA-Gly, murB, dapB, omlA genes and partial fur gene.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
3914642_f2_80	1636	6858	300	903		

Protein name	Locus Name	Acc#
<u>Description</u>		

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
3943802_f2_66	1637	6859	133	402	343	4.0e-31
Protein name			Locus Name		Acc#	
YjgF			gp:AF095578		AF095578	
Description						
Salmonella typhimurium YjgF (yjgF) gene, complete cds; and unknowngene.						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
3944687_f3_143	1638	6860	294	885	221	3.3e-18
Protein name			Locus Name		Acc#	
hypothetical protein AF0417			pir:A69302		A69302	
Description						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4064000_f1_29	1639	6861	88	267	83	0.0014
Protein name			Locus Name		Acc#	
probable integral membrane protein			pir:T37050		T37050	
Description						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4101561_c2_277	1640	6862	392	1179		
Protein name			Locus Name		Acc#	
Description						
NO-HIT						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4104636_c1_192	1641	6863	191	576	199	7.2e-16
Protein name			Locus Name		Acc#	
serine acetyltransferase			pir:G72349		G72349	
Description						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4149067_c3_341	1642	6864	194	585	230	3.7e-19
Protein name			Locus Name		Acc#	
serine acetyltransferase			pir:G72349		G72349	
Description						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4330032_f2_73	1643	6865	536	1611	135	4.6e-05
Protein name			Locus Name		Acc#	
			sp:Y143_SYNY3		P74442	
Description						

HYPOTHETICAL WD-REPEAT PROTEIN SLR0143

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4690675_f3_152	1644	6866	225	678		
Protein name			Locus Name		Acc#	
Description						

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4710902_f1_17	1645	6867	813	2442	756	5.2e-74
Protein name			Locus Name		Acc#	
			sp:BACA_BACLI		O68006	
Description						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4773261_c1_196	1646	6868	237	714	103	0.017

Protein name

Locus Name

Acc#

sp:YJBH_ECOLI

P32689

Description

PRECURSOR

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4802168_c3_406	1647	6869	206	621		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4882192_c1_194	1648	6870	409	1230	162	1.4e-09

Protein name

Locus Name

Acc#

probable lipopolysaccharide
N-acetylglucosaminyltransferase, rfbU

pir:F64500

F64500

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4898450_f2_62	1649	6871	255	768	458	2.6e-43

Protein name

Locus Name

Acc#

phnP protein (phnP) homolog

pir:D70166

D70166

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4954380_f2_90	1650	6872	620	1863	357	4.4e-47

Protein name

Locus Name

Acc#

oxaloacetate decarboxylase, subunit alpha
(oadA) homolog

pir:C69406

C69406

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
4957965_f3_130	1651	6873	265	798	300	1.4e-26

Protein name

Locus Name

Acc#

sp:DP3B_PSEPU

P13455

Description

DNA POLYMERASE III, BETA CHAIN,

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
5086537_f2_76	1652	6874	282	849	251	8.3e-21

Protein name

Locus Name

Acc#

putative histidine protein kinase

gp:REU82564

U82564

Description

hydrogenase-like protein small subunit(hoxB) gene, hydrogenase-like protein large subunit (hoxC) gene, and putative histidine protein kinase (hoxJ) gene, complete cds, and nickel permease (hoxN) gene, partial cds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
5260957_c1_193	1653	6875	440	1323	108	2.8e-05

Protein name

Locus Name

Acc#

sp:FER_METBA

P00202

Description

FERREDOXIN

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

5272656_c2_275	1654	6876	332	999	443	1.0e-41
----------------	------	------	-----	-----	-----	---------

Protein name	Locus Name	Acc#
--------------	------------	------

ss-1,4-galactosyltransferase	gp:SPCPS14E	X85787
------------------------------	-------------	--------

Description

S.pneumoniae cps14 locus.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

6048452_f1_21	1655	6877	78	237		
---------------	------	------	----	-----	--	--

Protein name	Locus Name	Acc#
--------------	------------	------

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

6414677_c1_188.....	1656	6878	405	1218	149	1.5e-07
---------------------	------	------	-----	------	-----	---------

Protein name	Locus Name	Acc#
--------------	------------	------

NADH dehydrogenase (ubiquinone), , 39 kDa subunit homolog	pir:H69478	H69478
---	------------	--------

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

6694425_c3_365.....	1657	6879	307	924	420	2.7e-39
---------------------	------	------	-----	-----	-----	---------

Protein name	Locus Name	Acc#
--------------	------------	------

hypothetical protein sll0744	pir:S77079	S77079
------------------------------	------------	--------

Description

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
6742762_f1_28	1658	6880	61	186		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
6834387_f1_19	1659	6881	120	363		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
6845277_c2_288	1660	6882	64	195		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
819433_c3_340	1661	6883	371	1116	339	1.0e-30

Protein name

Locus Name

Acc#

capsular polysaccharide biosynthesis homolog
yveT

pir:A70037

A70037

Description

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
-----------------	-------------	-------------	------------------	------------------	--------------	--------------------

960825_f2_69	1662	6884	465	1398	771	1.7e-76
--------------	------	------	-----	------	-----	---------

Protein name

Locus Name

Acc#

phosphate starvation inducible protein
homolog ylaK

pir:A69873

A69873

Description

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
-----------------	-------------	-------------	------------------	------------------	--------------	--------------------

9773281_f2_125	1663	6885	70	210		
----------------	------	------	----	-----	--	--

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
-----------------	-------------	-------------	------------------	------------------	--------------	--------------------

10573830_c1_296	1664	6886	168	507		
-----------------	------	------	-----	-----	--	--

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
-----------------	-------------	-------------	------------------	------------------	--------------	--------------------

10664128_c3_491	1665	6887	214	645	125	1.8e-11
-----------------	------	------	-----	-----	-----	---------

Protein name

Locus Name

Acc#

ribonuclease H, 1

pir:JC5787

JC5787

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

10720337_f1_75	1666	6888	211	636	258	4.0e-22
----------------	------	------	-----	-----	-----	---------

Protein name

Locus Name

Acc#

sp:YC08_YEAST

P37261

Description

HYPOTHETICAL 21.1 KD PROTEIN IN FUS1-AGP1 INTERGENIC REGION

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

10986288_f3_169	1667	6889	406	1221	149	2.2e-07
-----------------	------	------	-----	------	-----	---------

Protein name

Locus Name

Acc#

hypothetical protein BBI16

pir:G70241

G70241

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

10988261_f1_7	1668	6890	714	2145	1016	1.9e-102
---------------	------	------	-----	------	------	----------

Protein name

Locus Name

Acc#

DNA topoisomerase III topB

pir:H69724

H69724

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

11910250_f2_102	1669	6891	159	480		
-----------------	------	------	-----	-----	--	--

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
----------	------	------	-----------	-----------	-------	-------------

11924205_c3_539	1670	6892	110	333		
-----------------	------	------	-----	-----	--	--

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
1212762_c2_390	1671	6893	300	903	354	2.7e-32
Protein name			Locus Name		Acc#	
			gp:AB012957		AB012957	

Description

Vibrio cholerae genes for o-antigen synthesis, strain O22, completecds.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
12298425_c2_391	1672	6894	300	903	219	5.5e-18
Protein name			Locus Name		Acc#	
putative glycosyl transferase			gp:AF048749		AF048749	

Description

Bacteroides fragilis capsular polysaccharide biosynthesis operon, complete sequence.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
12501087_f2_116.....	1673	6895	168	507		
Protein name			Locus Name		Acc#	

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
12691280_c3_510.....	1674	6896	1022	3069	726	1.3e-81
Protein name			Locus Name		Acc#	
probable swf/snf helicase			pir:E71481		E71481	

Description

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
13885212_f3_220	1675	6897	388	1167		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
14116635_f2_158	1676	6898	416	1251		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
14736262_c3_545	1677	6899	427	1284	241	7.6e-20

Protein name

Locus Name

Acc#

MocB (Tn4399)

pir:B48487

B48487

Description

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
15017287_f3_252	1678	6900	837	2514	391	2.2e-32

Protein name

Locus Name

Acc#

enhanced entry protein EnhC

gp:AF057704

AF057704

Description

Legionella pneumophila EnhA (enhA), EnhB (enhB), and enhanced entryprotein EnhC (enhC) genes, complete cds.

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
16135886_f3_193	1679	6901	103	312		

Protein name Locus Name Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
16229025_f2_124.....	1680	6902	280	843	105	0.00081

Protein name Locus Name Acc#

sp:YS21_BORBU

Description

HYPOTHETICAL PROTEIN BBD21

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
16829717_f3_188.....	1681	6903	1951	5856	1011	1.5e-118

Protein name Locus Name Acc#

gp:AB016260

Description

Agrobacterium tumefaciens plasmid pTi-SAKURA, complete sequence.

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
16832885_f2_168.....	1682	6904	450	1353	1713	2.6e-176

Protein name Locus Name Acc#

hypothetical protein

pir:JQ1020

JQ1020

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
19562660_f2_134	1683	6905	478	1437	113	0.0074
Protein name			Locus Name		Acc#	
ES/130			gp:AF006751		AF006751	
Description						
Homo sapiens ES/130 mRNA, complete cds.						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
197131_f1_35	1684	6906	335	1008		
Protein name			Locus Name		Acc#	
Description						
NO-HIT						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
19945402_f1_57	1685	6907	375	1128		
Protein name			Locus Name		Acc#	
Description						
NO-HIT						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
1995941_f2_89	1686	6908	427	1284	1049	6.1e-106
Protein name			Locus Name		Acc#	
transposase			gp:AF038866		AF038866	
Description						
Bacteroides fragilis transposon Tn5520 transposase (bipH) and mobilization protein BmpH (bmpH) genes, complete cds.						

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
20213132_c2_406	1687	6909	61	186		

Protein name Locus Name Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
2036268_c3_584	1688	6910	84	255		

Protein name Locus Name Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
2053887_f2_143	1689	6911	114	345		

Protein name Locus Name Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
214526_c3_517	1690	6912	66	201		

Protein name Locus Name Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
21484662_f1_6	1691	6913	532	1599	144	9.4e-07
Protein name			Locus Name		Acc#	
			sp:M49_STRPY		P16947	
Description						
M PROTEIN, SEROTYPE 49 PRECURSOR						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
21515632_f1_17	1692	6914	785	2358	3809	0.0
Protein name			Locus Name		Acc#	
tetracycline resistance element regulator RteA			pir:A41860		A41860	
Description						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
21526437_f2_130.....	1693	6915	138	417		
Protein name			Locus Name		Acc#	
Description						
NO-HIT						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
21601625_c1_295.....	1694	6916	133	402		
Protein name			Locus Name		Acc#	
Description						
NO-HIT						

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
21679626_f1_45	1695	6917	174	525		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
22459443_c3_486.....	1696	6918	150	453	130	1.5e-08

Protein name

Locus Name

Acc#

gp:APU72238

U72238

Description

Anabaena PCC7120 ORFR1, ORFR2, ORFR3, ORFR4, and ORFR5 genes, complete sequences.

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
22687625_c3_523.....	1697	6919	433	1302	95	4.3e-05

Protein name

Locus Name

Acc#

phage abortive infection protein

pir:T30326

T30326

Description

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
22689031_c2_386.....	1698	6920	386	1161	1201	4.7e-122

Protein name

Locus Name

Acc#

UDP-galactopyranose mutase

gp:SPAJ6986

AJ006986

Description

Streptococcus pneumoniae type 33F DNA, capsular gene cluster.

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
22774087_f2_126	1699	6921	246	741	85	0.0034
<u>Protein name</u>			<u>Locus Name</u>		<u>Acc#</u>	
non-structural 5a protein			gp:HCU56570		U56570	
<u>Description</u>						
Hepatitis C virus isolate 925821 non-structural 5a (NS5a) gene, partial cds.						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
22860128_f3_258	1700	6922	83	252	64	0.031
<u>Protein name</u>			<u>Locus Name</u>		<u>Acc#</u>	
			sp:SPRC_XENLA		P36378	
<u>Description</u>						
(OSTEONECTIN) (ON) (BASEMENT MEMBRANE PROTEIN BM-40)						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
22892905_f2_115.....	1701	6923	452	1359	2093	1.4e-216
<u>Protein name</u>			<u>Locus Name</u>		<u>Acc#</u>	
			gp:BNRRTEAB			
<u>Description</u>						
Bacteroides thetaiotaomicron rteA and rtaB genes involved in production of plasmid-like forms, complete cds, and tetQ gene, 3'end.						

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
23443750_f2_148.....	1702	6924	433	1302	160	2.4e-08
<u>Protein name</u>			<u>Locus Name</u>		<u>Acc#</u>	
actin binding protein MAYVEN			gp:AF059569		AF059569	
<u>Description</u>						
Homo sapiens actin binding protein MAYVEN mRNA, complete cds.						

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
23492786_f2_162	1703	6925	85	258		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
23596911_f2_139	1704	6926	77	234		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
23644552_f1_31	1705	6927	941	2826	537	8.3e-85

Protein name

Locus Name

Acc#

gp:BFU63096

U63096

Description

Bacteroides fragilis (bctA) gene, complete cds.

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
23710777_c1_327	1706	6928	146	441	81	0.029

Protein name

Locus Name

Acc#

hypothetical protein

gp:AF036485

Description

Plasmid pNZ4000, complete sequence.

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
24251937_f1_1	1707	6929	94	285		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
24259442_f3_246.....	1708	6930	313	942	370	5.4e-34

Protein name

Locus Name

Acc#

sp:GSPA_BACSU

P25148

Description

GENERAL STRESS PROTEIN A

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
24347090_f1_34.....	1709	6931	198	597		

Protein name

Locus Name

Acc#

Description

NO-HIT

<u>ORF Name</u>	<u>NTID</u>	<u>AAID</u>	<u>NT Length</u>	<u>AA Length</u>	<u>Score</u>	<u>Probability</u>
24410812_f1_58.....	1710	6932	269	810		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24415885_f3_171	1711	6933	80	243		

Protein name

Locus Name

Acc#

Description

NO-HIT

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24491255_c3_506.....	1712	6934	299	900	283	9.0e-25

Protein name

Locus Name

Acc#

ribonuclease III (rnc) homolog

pir:H70187

H70187

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24609762_c3_576.....	1713	6935	101	306	82	0.0018

Protein name

Locus Name

Acc#

vrl1 protein

pir:T17388

T17388

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24632687_f3_256.....	1714	6936	637	1914	1505	2.9e-154

Protein name

Locus Name

Acc#

arginine decarboxylase, 2:protein
slr0662:protein slr0662

pir:S76771

S76771

Description

ORF Name	NTID	AAID	NT Length	AA Length	Score	Probability
24640876_f3_248.....	1715	6937	488	1467	125	0.00018

Protein name

Locus Name

Acc#

complement C9 precursor

pir:C9HU

Description